MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Dec. 2007, p. 576–599 Vol. 71, No. 4
1092-2172/07/$08.00+0 doi:10.1128/MMBR.00015-07

Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs†

F. Robert Tabita,[1]* Thomas E. Hanson,[2] Huiying Li,[3] Sriram Satagopan,[1] Jaya Singh,[4] and Sum Chan[3]

Department of Microbiology and Plant Molecular Biology/Biotechnology Program, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210-1292[1]; Graduate College of Marine and Earth Studies, Delaware Biotechnology Institute, University of Delaware, 127 DBI, 15 Innovation Way, Newark, Delaware 19711[2]; Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Department of Chemistry and Biochemistry, University of California, Los Angeles, Box 951570, Los Angeles, California 90095-1570[3]; and Department of Plant Cellular and Molecular Biology, The Ohio State University, 582 Aronoff Laboratory, 318 W. 12th Avenue, Columbus, Ohio 43210-1292[4]

INTRODUCTION

Virtually all the organic carbon found on earth is derived from oxidized inorganic sources such as gaseous carbon dioxide and carbon monoxide as well as soluble and insoluble bicarbonate and carbonate deposits. These various forms of inorganic carbon are in chemical equilibrium on earth, and the relative concentration of each species in specific environments is dependent on localized parameters such as temperature, pH, and pressure. Moreover, it is widely believed that levels of anthropogenic CO_2 are steadily increasing in the earth's atmosphere, and predictions are that these levels will increase steadily, with consequent effects related to the potential warming of the earth. In order to surmount the rather considerable energy required to chemically convert oxidized inorganic carbon to reduced organic carbon on a global scale, living organisms and specific biological macromolecules eventually evolved to catalyze this process. Fortunately, terrestrial and marine plants and specialized microbes developed the ability to remove and assimilate considerable amounts of CO_2 from the atmosphere and, in the process, formed the necessary organic carbon skeletons required to sustain the biosphere. More specifically, different enzymatic schemes evolved to catalyze inorganic carbon reduction such that there are currently four known metabolic pathways by which organisms can grow using CO_2 as their sole source of carbon (28, 79). These include the Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway, the reductive tricarboxylic acid cycle, the Wood-Ljungdahl acetyl coenzyme A pathway, and the hydroxypropionate pathway. From a biogeochemical standpoint, the CBB reductive pentose phosphate pathway is by far the major means by which CO_2 is reduced to form organic carbon. In this scheme, the sugar bisphosphate ribulose-1,5-bisphosphate (RuBP) serves as the acceptor molecule for CO_2, with the enzyme RuBP carboxylase/oxygenase (RubisCO) catalyzing the actual primary CO_2 fixation reaction. RubisCO is found in most autotrophic organisms, ranging from diverse prokaryotes, including photosynthetic and chemolithoautotrophic bacteria and archaea, to eukaryotic algae and higher plants. RubisCO is also clearly the most abundant protein found on earth (21), as it can comprise up to 50% of the total soluble protein found in leaf tissue or within specific microbes (67, 68). Such exaggerated abundance is most likely due to the poor catalytic efficiency of RubisCO, with a turnover number ($\sim 5\ s^{-1}$) that is among the lowest for any biological catalyst (13, 68).

* Corresponding author. Mailing address: Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210-1292. Phone: (614) 292-4297. Fax: (614) 292-6337. E-mail: Tabita.1@osu.edu.

† Supplemental material for this article may be found at http://mmbr.asm.org.

DIFFERENT MOLECULAR FORMS FOR THE SAME (AND DIFFERENT) FUNCTIONS

Classically, RubisCO is comprised of both large (catalytic) and small subunits to form a massive hexadecameric protein structure with an M_r of about 550,000, i.e., eight copies of both large ($\sim$ 55,000 M_r) and small ($\sim$ 15,000 M_r) polypeptides in an $(L_2)_4(S_4)_2$ structure (4, 35). This type of enzyme, now called form I, is the predominant RubisCO form found in nature, and it is present in terrestrial and marine plants, eukaryotic algae, cyanobacteria, and most phototrophic and chemolithoautotrophic proteobacteria (68). The name form I was originally used to distinguish this type of RubisCO from another structurally simpler form of the enzyme that was shown to be a dimer of only large subunits, which was discovered originally in the nonsulfur phototrophic bacterium *Rhodospirillum rubrum* (69, 70). Interestingly, another nonsulfur purple phototrophic bacterium, *Rhodobacter sphaeroides*, also appeared to contain this second structural form of RubisCO (albeit in higher aggregates of large subunits) and was originally isolated as a second peak of activity after ion-exchange fractionation of extracts from induced *R. sphaeroides*. Form I RubisCO was isolated from the same crude extracts, i.e., in the first activity peak that eluted from the column (29). Thus, the enzyme from the second activity peak (peak II), which contained the novel structural form analogous to *R. rubrum* RubisCO, was eventually called the form II enzyme to distinguish it from the first peak of activity or the form I enzyme. Form II RubisCO proteins were shown to catalyze the same reaction as form I RubisCO, and both enzymes catalyze an oxygen fixation reaction whereby the enediol of RuBP is attacked by molecular oxygen. The form II enzyme, comprised only of multimers of large-type subunits $[(L_2)_x]$, shows only about 30% amino acid sequence identity to form I large subunits. In addition, form II enzymes all appear to be less efficient in partitioning the two gaseous substrates of RubisCO, CO_2 and O_2. Most importantly, the form II enzyme takes on a distinct physiological role, as it is used primarily to enable the CBB pathway to balance the redox potential of the cell under select growth conditions (19, 68, 74). To this day, the relative differences and similarities in primary structure serve as a convenient means to classify all the different forms of RubisCO found in nature.

By the mid-1990s, it was recognized that the form I enzyme could be further classified, according to amino acid sequence homologies, as either "green" (cyanobacterial, algal, and plant) and "red" (phototrophic bacterial and nongreen eukaryotic algal) (16, 19, 67, 68, 74). As more RubisCO gene sequences became available, the green enzymes were further subdivided into forms IA and IB, and the red enzymes were subdivided into forms IC and ID (67, 68) (Fig. 1). Form II bacterial enzymes, and even eukaryotic homologs found in symbiotic dinoflagellates, all appear to be fairly closely related, and there is no clear subdivision. This convenient division into different phylogenetic and catalytically distinct structural forms (forms I and II) lasted for about 20 years. The more recent explosion of complete genomic sequencing projects has led to putative RubisCO sequences showing up in some unusual places, including organisms that use alternatives to the CBB pathway to fix CO_2 and even microorganisms that do not use CO_2 as a major carbon source. For example, it was shown that various archaea, including those that use other means for primary CO_2 assimilation or those that may even grow on organic compounds, contain genes that encode a bona fide functional RubisCO (25, 73; F. R. Tabita, G. M. Watson, and J. P. Yu, presented at the 98th Meeting of the American Society for Microbiology, 1998). Moreover, phylogenetic analyses clearly placed the archaeal RubisCO sequences in a separate category, which was termed form III (68, 73). Thus, by the late 1990s, it was apparent that nature still had some surprises for RubisCO biochemists and evolutionists (RubisCOlogists), and the rather comfortable and long-standing classification of RubisCO into only forms I and II was clearly and obviously incomplete and actually incorrect. For those interested in structure-function relationships, the advent of the form III enzymes, obtained from organisms that never see molecular oxygen, offers tantalizing possibilities to learn more about how the active site of RubisCO might have evolved. This is especially relevant since it was found that several archaeal enzymes are highly sensitive to molecular oxygen and have extremely poor capabilities to discriminate between CO_2 and O_2 (27, 37, 73) due in part to an extremely high affinity of these enzymes for O_2 (37).

The RubisCO-Like Protein (Form IV), a Homolog of RubisCO

It was first noted in 1999 (68), via just-completed genomic sequencing projects, that the green sulfur phototrophic bacterium *Chlorobium tepidum* and the heterotroph *Bacillus subtilis* contained putative RubisCO genes that were clearly not of the form I and form II types. These RubisCO genes were initially thought to be in the newly discovered form III archaeal class since those sequences that were not of form I and form II all seemed to be quite different from each other as well as from the established form I and form II sequences (68). Interestingly, *C. tepidum*, though autotrophic, does not assimilate CO_2 via the CBB pathway to obtain organic carbon, and *B. subtilis* does not use CO_2 as a carbon source at all. Subsequent analyses showed that the putative RubisCO genes from these organisms were distinct from bona fide form III RubisCOs from archaea, as the *C. tepidum* and *B. subtilis* sequences both contain dissimilar residues at positions analogous to the mechanistically significant residues that are important for catalysis in RubisCO counterparts, and the purified recombinant *C. tepidum* protein was unable to catalyze RuBP-dependent carboxylation (31). Moreover, disruption of the gene in *C. tepidum* resulted in sulfur deposition into the surrounding media as well as distinct effects on autotrophic growth. Based on these studies and the fact that this protein resembles bona fide RubisCOs (about 35% identity at the amino acid level), the RubisCO homolog from *C. tepidum* was termed the RubisCO-like protein (RLP) and categorized as form IV RubisCO (31). Further studies confirmed the role of *C. tepidum* RLP in sulfur metabolism (thiosulfate oxidation), and its disruption led to a general stress response (30). As for *B. subtilis*, genetic studies (45, 63), followed by biochemical analyses (7), showed that its RLP (or YkrW/MtnW) participates in a methionine salvage pathway and catalyzes the enolization of the RuBP analog 2,3-diketo-5-methylthiopentyl-1-P. Based on phylogenetic analyses of currently available RLP sequences (see below), there appear to be six different clades of RLP or form IV RubisCO (Fig. 1); the

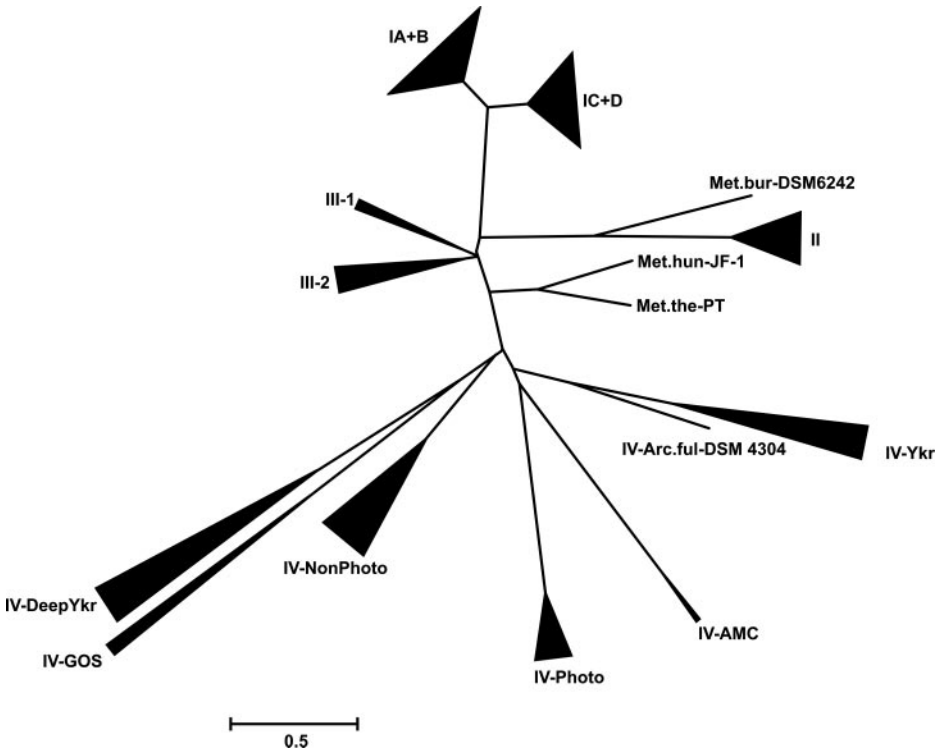


FIG. 1. Unrooted NJ tree of RubisCO/RLP lineages. To construct this tree, a total of 193 sequences were aligned with MEGA 3.1 (38) and evaluated by ProtTest (1), and the tree was then constructed using the equal-input model with a gamma rate distribution of 1.554. The total numbers of sequences considered in each lineage were 35 for I-A, 16 for I-B, 9 for I-C, 22 for I-D, 20 for II, 10 for III-1, 4 for III-2, 20 for IV-NonPhoto, 2 for IV-EnvOnly, 14 for IV-Photo, 16 for IV-DeepYkrW, 12 for IV-YkrW, and 5 for IV-GOS. The width of the arrows is directly proportional to the number of sequences considered for each clade. For a complete list of sequences and sources, see Table S1 in the supplemental material. The scale bar represents a difference of 0.5 substitutions per site. Bootstrap values for nodes are shown in Fig. 2A. Single-sequence abbreviations and sequence identifiers are as follows: IV-Arc.ful-DSM 4304, *Archaeoglobus fulgidus* strain DSM4304 (GenBank accession number NP_070416); Met.bur-DSM6242, *Methanococcoides burtonii* strain DSM6242 (accession number ZP_00563653); Met.hun-JF-1, *Methanospirillum hungatei* strain JF-1 (accession number YP_503739); Met.the-PT, *Methanosaeta thermophila* strain PT (accession number ZP_01153096).

function(s) of these proteins, as currently understood, will be further discussed below.

The RubisCO Superfamily at Present

For years, RubisCO has been one of the most deeply sequenced protein families. However, up until about 10 years ago, except for a few microbial genes, most of the known RubisCO gene sequences were obtained from different plants, all of which were shown to be closely related. These earlier sequencing projects were all directed towards the obvious interest in this protein as a target for crop improvement. However, the recent profusion of microbial genome sequencing projects from truly diverse organisms plus the burgeoning documentation of metagenomic RubisCO sequences from environmental samples (2, 22, 23, 34, 35, 47, 55, 65, 75, 78) have opened up new vistas and suggest that meaningful evolutionary analysis of this protein may now be undertaken and proceed towards an informed conclusion. This analysis is feasible despite the rather obvious misannotation of various RubisCO sequences in the NCBI protein database as methionine sulfoxide reductase A (see GenBank accession NP_248230 for one example from *Methanocaldococcus jannaschii*).

Phylogenetic analyses of RubisCO and RLP sequences indicate that there are at least three distinct lineages of bona fide RubisCO and six distinct clades of RLP molecules (Fig. 1). The well-studied form I and form II groups are each monophyletic and, despite their clear separation, are somewhat related to each other. Form III sequences are recognizably distinct from forms I and II by any phylogenetic reconstruction method employed (31) (Fig. 2), which initially suggested a relationship to RLP (68). However, all form III proteins analyzed thus far can catalyze RubisCO activity (27, 73) in vitro, while no RLP has ever been documented to catalyze RuBP-dependent CO_2 fixation, undoubtedly due to the absence of critical conserved active-site residues (13, 68) in the latter (Fig. 3). Thus, the only currently known bona fide RubisCO sequences are those found within forms I, II, and III. Outlying sequences observed in recently sequenced methanogen genomes will be discussed below.

The six remaining clades in the RubisCO form IV (RLP) lineage have been termed IV-Photo (found in phototrophic bacteria), IV-NonPhoto (found in nonphototrophic bacteria), IV-AMC (acid mine consortia), IV-YkrW, IV-DeepYkr, and IV-GOS (global ocean sequencing sequencing program) based on characteristics of the source organisms, prior designation of the gene product, and/or relationship to other sequences (Fig.

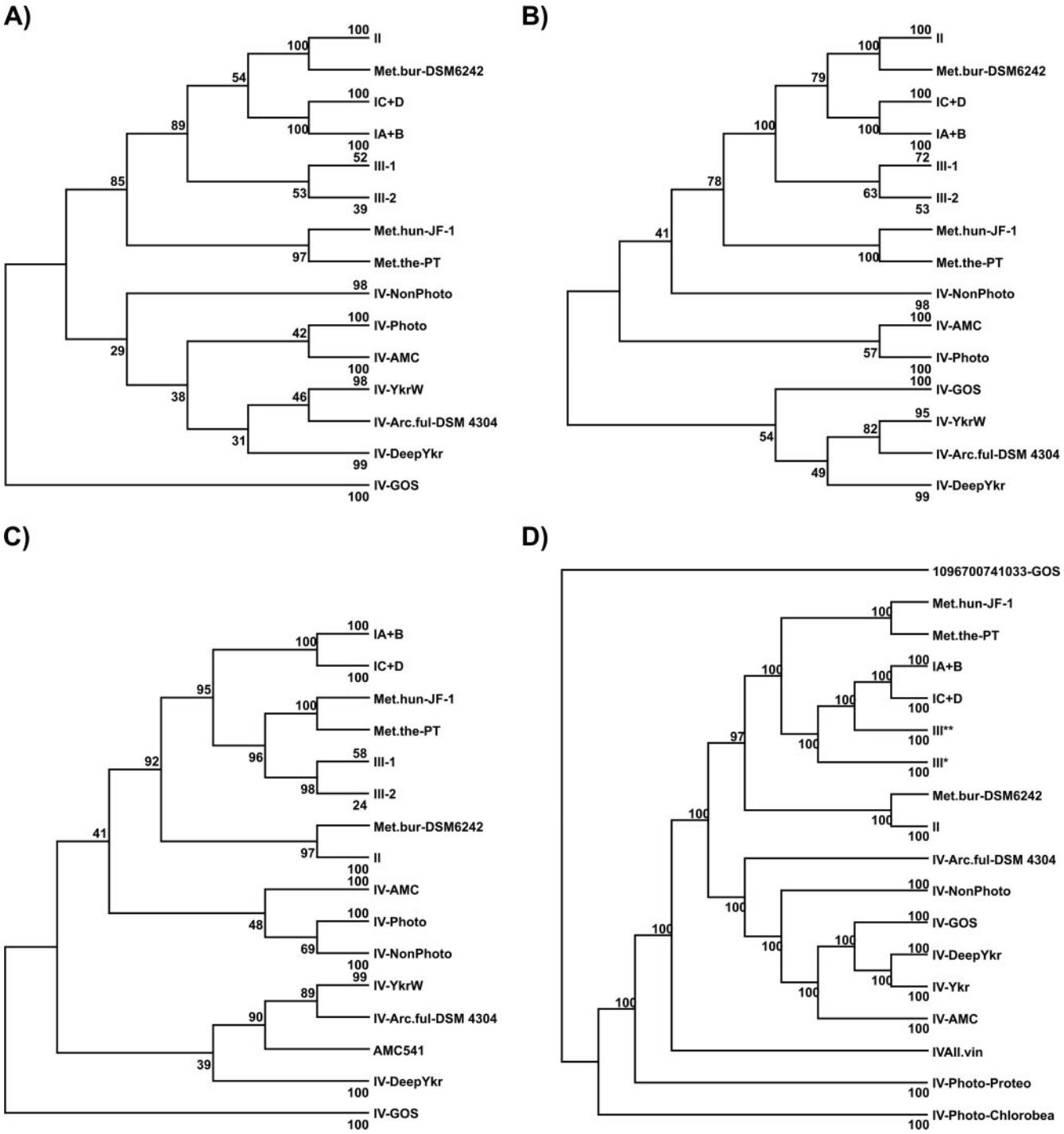


FIG. 2. Comparison of RubisCO/RLP tree topologies reconstructed with NJ (A), ME (B), UPGMA (C), and MP (D). All except MP assumed a distribution of 1.554 of evolutionary rates across four categories as calculated by ProtTest (1). Values at nodes represent bootstrap support observed in 1,000 trials per method. IV-Arc.ful-DSM 4304, *Archaeoglobus fulgidus* strain DSM4304 (GenBank accession number NP_070416); Met.bur-DSM6242, *Methanococcoides burtonii* strain DSM6242 (accession number ZP_00563653); Met.hun-JF-1, *Methanospirillum hungatei* strain JF-1 (accession number YP_503739); Met.the-PT, *Methanosaeta thermophila* strain PT (accession number ZP_01153096).

1). All sequences in the RLP lineage satisfy two major criteria, namely, that they fail to cluster with any bona fide RubisCO sequence in phylogenetic trees and that each sequence contains nonconservative substitutions at the positions normally occupied by conserved RubisCO active-site residues, rendering these proteins incapable of RubisCO catalysis (Fig. 3). While activity screening has been limited to only a few such recombinant proteins, e.g., *C. tepidum* RLP (31), *Rhodopseudomonas palustris* RLP1 and RLP2 (S. Romagnoli and F. R. Tabita, unpublished results), *Rhodospirillum rubrum* RLP (J. Singh

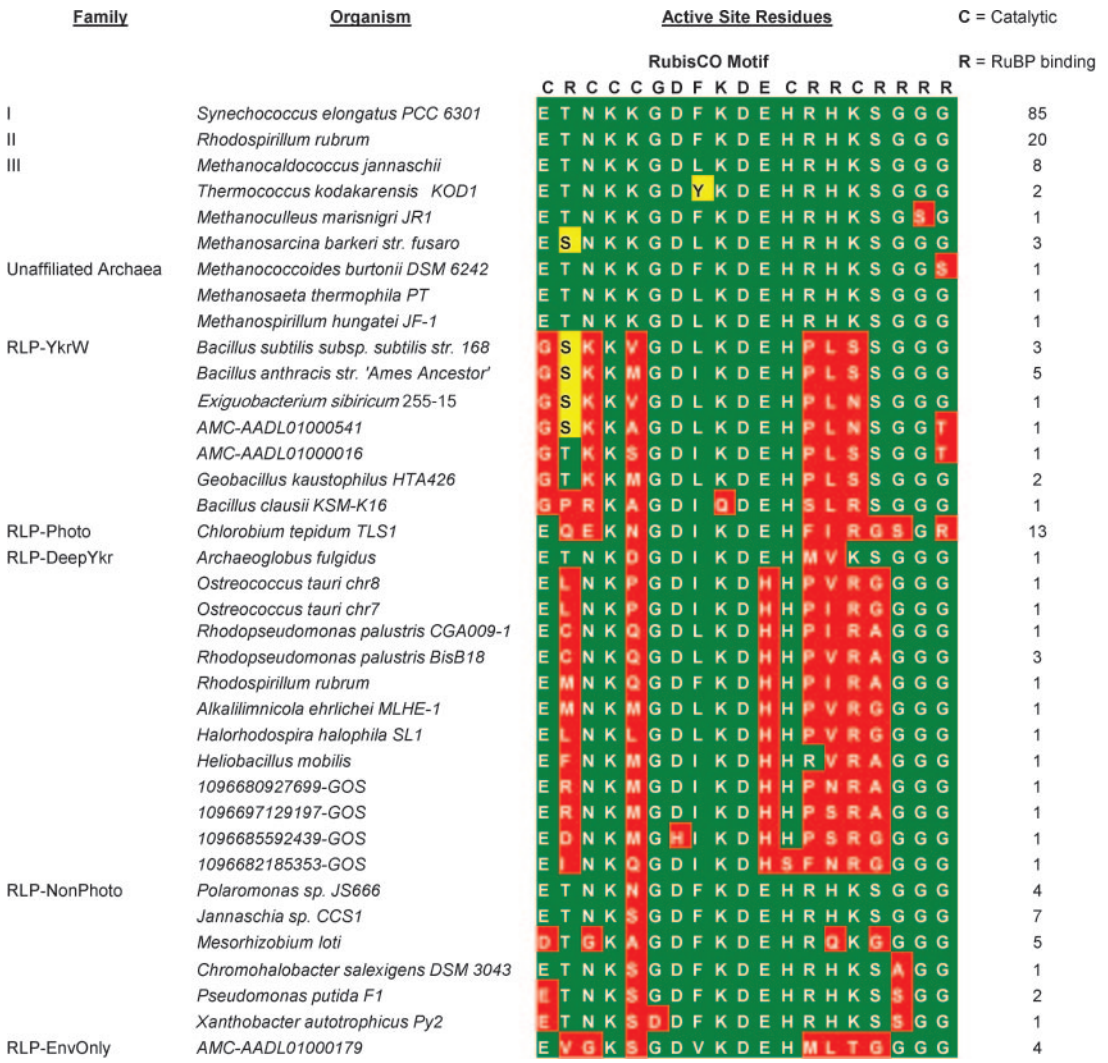


FIG. 3. Conservation of RubisCO active-site residues in RubisCO/RLP family members as noted previously by Cleland et al. (13) and Tabita (68). All form III RubisCO and RLP (form IV) sequences used in the reconstruction of phylogenetic relationships are included. Residues are noted in single-letter IUPAC code. Positions shaded green indicate conservation, while yellow indicates a semiconservative substitution and red indicates a nonconservative substitution. C, catalytic residue; R, RuBP binding residue.

and F. R. Tabita, unpublished results), and *B. subtilis* (8) and *Geobacillus kaustophilus* YkrW proteins (33), it is likely that all members of these families lack RuBP carboxylase or oxygenase activity. Moreover, as discussed elsewhere in this review, structural and functional evidences indicate that the *C. tepidum* RLP is incapable of productively binding the transition-state analog 2-carboxyarabitinol-1,5-bisphosphate (CABP) (31, 39; S. Satagopan and F. R. Tabita, unpublished results). Comparative tree topologies obtained by different phylogenetic inference methods (Fig. 2) are discussed below.

Sequence Conservation in the RubisCO Superfamily

Overall sequence conservation between lineages in the RubisCO large-subunit superfamily is detectable with an average of 31% amino acid sequence identity across the 193 nonredundant, full-length sequences analyzed (see Table S1 in the supplemental material). This sequence set included all available full-length RLP and metagenomic RubisCO amino acid sequences present in public databases as of April 2007, including those that recently became available from the global ocean-sampling (GOS) expedition (80). The minimum sequence identity observed between any two sequences in this data set was 6.9%, which is between a IV-DeepYkr sequence from the marine chlorophyte *Ostreococcus tauri* and a metagenomic form I sequence from the GOS sequence collection. Sequence conservation within a given lineage is variable but significantly higher than the average, ranging from an 85% mean in-group identity in the form IB lineage to 49% in the form IV-Non-Photo lineage (Table 1). The two exceptions to this rule are the form IV-DeepYkr lineage, which shares only 37% average in-group sequence identity, and the IV-GOS clade, which shares an average of 43% in-group sequence identity. The low sequence conservation and wide size range observed in both groups suggest that they may contain collections of single representatives of rarely observed RLP lineages. There is also

TABLE 1. RubisCO and RLP lineage properties and phylogenetic distribution[a]

Lineage	Protein size (aa)	% Identity		Phylogenetic distribution
		Avg	Min	
I-A	470–479	79	43	*Alpha*-, *Beta*-, and *Gammaproteobacteria*, *Cyanobacteria*, *Prochlorales*, Sargasso Sea metagenome, GOS metagenome
I-B	470–477	85	79	*Cyanobacteria*, *Prochlorales*, eukaryotes-*Viridiplantae* (Streptophyta, Chlorophyta), Euglenozoa, Sargasso Sea metagenome
I-C	477–488	79	67	*Alpha*- and *Betaproteobacteria*, chloroflexi
I-D	459–490	78	51	*Alpha*-, *Beta*-, and *Gammaproteobacteria*, eukaryotes-stramenopiles, Rhodophyta, *Haptophyceae*
II	458–585	68	44	*Alpha*-, *Beta*-, and *Gammaproteobacteria*, eukaryotes-Alveolata (*Dinophyceae*)
III-1	414–444	68	48	Methanogenic crenarchaeota
III-2	425–428	53	41	Methanogenic and thermophilic crenarchaeota, thermophilic and halophilic euryarchaeota
IV-NonPhoto	392–432	49	38	*Alpha*-, *Beta*-, and *Gammaproteobacteria*, chloroflexi
IV-DeepYkr	368–604	37	20	*Alphaproteobacteria*, clostridia, nonmethanogenic euryarchaeota, eukaryotes-*Ostreococcus tauri*
IV-AMC	307–393	64	51	Acid mine drainage microbial consortium
IV-GOS	363–417	43	27	GOS sequence collection
IV-Photo	428–457	70	58	*Alpha*- and *Gammaproteobacteria*, chlorobia
IV-YkrW	374–414	58	24	*Firmicutes*, acid mine drainage microbial consortium

[a] The average percent identity within a lineage and the minimum (Min) identity observed within a lineage were calculated by pairwise comparison of all lineage members. Lineages were defined on the basis of the NJ tree shown in Fig. 1. aa, amino acids.

relatively low concerted covariation of active-site residues in the IV-DeepYkr group, further suggesting that it may require further subdivision as more sequences become available. Furthermore, recent findings, based on mass spectrometry-based discrimination of expressed protein products, suggest that recombination of diverse genomes from acidophilic bacteria has occurred in an acid mine environment. Such genomic recombination apparently resulted in the creation of a chimeric RLP with potentially novel functions among the acidophiles of this environment (41). With respect to RLP, it will be interesting to determine how widespread such recombination events are and their physiological consequences.

Only four residues are absolutely conserved among all members of the RubisCO superfamily when they are aligned on the basis of sequence similarity, with no specific consideration of structural motifs. These residues are Gly-122/110/100, Lys-175/166/153, Asp-203/193/191, and Gly-322/316/297 in representative enzymes of form I, form II, and form III from *Spinacia oleracea*, *Rhodospirillum rubrum*, and *Methanocaldococcus jannaschii*, respectively. Relaxing this conservation requirement to 99% of the sequences analyzed results in the identification of 10 additional residues. Three of these highly conserved residues, Asp-198/188/176, Lys-201/191/179, and Asp-203/193/181, lie within the "RubisCO motif." Lys-201/191/179 is the residue that becomes carbamylated when RubisCO is "activated" by CO_2 in the presence of a divalent metal prior to the actual catalytic event (see Fig. 8), i.e., when the RubisCO-CO_2-Me^{2+} ternary complex is attacked by a second molecule of CO_2 or O_2. Lys-175/166/153 is involved in the initial deprotonation and final protonation steps of the catalytic cycle. Asp-203/193/181 is one of the key metal binding ligands along with Glu-204/194/182, which is conserved at 90% identity among all RubisCO and RLP sequences. The highly conserved glycines have not been ascribed specific roles in RubisCO structure or function. If the stringency for conservation is relaxed to include those positions where there is a 90% consensus among all sequences, a total of 25 residues may be identified (Table 2). By comparing these 25 residues with those previously assigned functions by mutagenesis or structural studies, the conserved functions among all RubisCO large-subunit sequences appear to be Mg^{2+} binding, acid-base chemistry, substrate hydration, and a partial P1 binding site. The additional 18 conserved but non-active-site residues may well reflect unrecognized players in catalysis or protein stability or keystone residues critical to establishing or maintaining the structure of the active enzyme. The fact that the overall monomer structures of all RubisCO large-subunit superfamily members are quite similar supports the notion that there may be a conserved set of residues that are critical for folding and maintaining this general structure. However, underlying all these structural comparisons is the realization that while the authentic RubisCO proteins (forms I, II, and III) all catalyze the same reactions, many of these proteins, even those from the same clade, may have widely different enzymatic properties, especially the ability to discriminate between CO_2 and O_2 and perhaps other kinetic properties as well. This functional diversity is particularly evident among the closely related form IC enzymes (68). Indeed, even proteins whose structures are virtually superimposable, with up to nearly 90% sequence identity, may possess vastly different kinetic properties. Thus, while structural (discussed below) and sequence comparisons offer interesting insights into potential functional alterations, it is very often difficult to predict the enzymatic properties of individual RubisCO proteins.

Evidence for Distinct Functions among RLP Lineages

(i) Active-site substitution patterns and implications from functional studies. Distinct histories for each RLP lineage are supported by a common pattern of active-site substitutions observed within a given lineage that is not shared with other lineages (Fig. 3). These common patterns of substitution are expected to affect the functionality of these enzymes

TABLE 2. Residues conserved at various percentages across all RubisCO/RLP sequences analyzed

Amino acid	Position			Conservation at:			Description
	Spinacia oleracea form I RubisCO	*Rhodospirillum rubrum* form II RubisCO	*Methanocaldococcus jannaschii* form III RubisCO	90%	95%	100%	
Gly	122	110	100	X[a]	X	X	
Asp	137	125	115	X			
Pro	141	129	119	X			
Gly	150	138	128	X	X		
Pro	151	139	129	X	X		
Lys	175	166	153	X	X	X	Proton acceptor/donor
Pro	176	167	154	X			
Gly	179	170	157	X	X		
Gly	195	186	173	X			
Gly	196	187	174	X	X		
Asp	198	188	176	X	X		
Lys	201	191	179	X	X		Carbamate, Mg ligand
Asp	203	193	181	X	X	X	Mg ligand
Glu	204	194	182	X			Mg ligand
Gly	233	223	211	X	X		
His	294	287	269	X	X		General base easing water attack at C3
Gly	308	302	283	X			
Arg	319	313	294	X	X		
Gly	322	316	297	X	X	X	
Gly	381	370	355	X			
Gly	395	384	369	X	X		
Gly	403	393	377	X	X		P1 phosphate binding
Gly	404	394	378	X			
Gly	405	395	379	X			P1 phosphate binding
His	409	399	383	X			
Gly	416	406	390	X			

[a] An "X" indicates that the residue is present at the given level of conservation.

given that certain features of the RubisCO active site are tightly conserved among all RLPs. This suggests that each lineage is a variation on a central structural or functional theme.

Given this variation, it seems unlikely that members of one lineage would functionally substitute for a member from another RLP family, although evidence exists otherwise (discussed below). Currently, detailed functional studies have been carried out for only four RLPs, *C. tepidum* RLP (30, 31), the YkrW/MtnW proteins of *Bacillus subtilis* and *Geobacillus kaustophilus* (8, 33, 45, 63), and the YkrW-like RLP from the cyanobacterium *Microcystis aeruginosa* (11). Thus far, the three-dimensional structures have been solved only for *C. tepidum* RLP (37), *R. palustris* RLP2 (this paper), and *G. kaustophilus* RLP (33, 39). The RLP from *C. tepidum* and the RLP2 from *R. palustris* are structurally very similar at the active site but possess four different active-site residues compared to the *B. subtilis* and *G. kaustophilus* proteins. Specific catalytic residues appear to be differentially conserved among the two lineages. The major difference is the Glu versus the Lys at Asn-123 (spinach RubisCO numbering), suggesting possible differences in hydrogen-bonding patterns with their respective substrates. In addition, Asn versus Val/Met identities at the Lys-177 position in *C. tepidum* versus *B. subtilis* groups of RLPs, respectively, may indicate different needs or participants for proton abstraction at the presumptive active site (see below), whereas Phe versus Pro identities at Arg-295, the residue that interacts with P2 phosphate in spinach RubisCO, likely indicate that each type of RLP reacts with distinct substrates with different hydrophobicities at the P2 site.

The *B. subtilis* YkrW/MtnW protein and, more recently, its *M. aeruginosa* and *G. kaustophilus* RLP homologs, have all been shown to function as a 2,3-diketo-5-methylthiopentyl-1-phosphate enolase in the methionine salvage pathway. Thus, a *B. subtilis* mutant lacking YkrW/MtnW has a relatively constrained phenotype that is manifested only under severe sulfur starvation conditions (45, 63). Based on structural comparisons discussed elsewhere in this review, it appears that 2,3-diketo-5-methylthiopentyl-1-phosphate is not compatible with the active-site pocket in *C. tepidum* RLP or *R. palustris* RLP2 (39). Thus, it was not surprising that inactivating these genes resulted in strains with distinct phenotypic properties in different organisms. For example, an insertionally inactivated RLP mutant of *C. tepidum* (strain Ω::RLP) had a highly pleiotropic phenotype, with defects observed in pigmentation, the ability to metabolize some sulfur compounds, and the aberrant expression of stress response proteins (31). More specifically, strain Ω::RLP is unable to oxidize thiosulfate efficiently, although the ability to oxidize sulfide remains unperturbed (30). Strain Ω::RLP is also deficient in oxidizing elemental sulfur, as it was found to produce significantly more extracellular elemental sulfur than the wild type (31).

A null mutation in the gene encoding RLP in *C. tepidum* also results in the overproduction of two oxidative stress response-related proteins, i.e., a thiol-specific antioxidant (Tsa) protein and superoxide dismutase. The levels of these two proteins are

12- and 3-fold enhanced, respectively, in the Ω::RLP strain compared with the wild type. The accumulation of these proteins correlates with the transcript levels of the corresponding genes (30). The Ω::RLP strain is also significantly more resistant to hydrogen peroxide exposure during growth than is the wild type (30). Further analyses indicate that the *C. tepidum* genome also encodes two potentially relevant transcriptional regulators, i.e., the ferric ion uptake regulator (Fur) and the peroxide regulator (PerR). Since these regulators are reported to be involved in the regulation of oxidative stress response genes in various bacteria including *Escherichia coli*, *Bacillus subtilis*, and *Staphylococcus aureus*, the possibility that RLP might be involved with the function of these regulators was considered. However, insertional inactivation of both the *fur* and *perR* genes of *C. tepidum* did not affect the accumulation of the Tsa and superoxide dismutase proteins in the Ω::RLP mutant strain (Singh and Tabita, unpublished).

How RLP specifically contributes to sulfur oxidation and oxidative stress in chlorobia is still unknown. These areas have received relatively little experimental attention in chlorobia to date, although this is beginning to change with the exploitation of available genomic data (12). Genes encoding RLPs have been found in all *Chlorobium* genomes sequenced to date (see http://img.jgi.doe.gov/cgi-bin/pub/main.cgi for details) even though these strains vary considerably in the spectra of reduced sulfur compounds used to support growth. Regarding oxidative stress, *Chlorobium* sp. strain GSB1, recently isolated from a hydrothermal vent sample (9), was found to maintain viability during prolonged exposure to molecular oxygen only in the absence of light and sulfide. Clearly, experiments utilizing oxidative stress elicitors other than molecular oxygen (i.e., organic hydroperoxides, methyl viologen, or diamide) in addition to experiments examining the interplay of light, sulfur compounds, and oxygen are required. Such studies of stress physiology and sulfur oxidation will likely contribute to delineating the function of RLP in *C. tepidum*.

Phylogenetic analyses (Fig. 1) indicated that some organisms (i.e., *Rhodopseudomonas palustris*, *Rhodospirillum rubrum*, and *Microcystis aeruginosa*) contain both bona fide RubisCO as well as RLP. Indeed, *R. palustris*, a purple nonsulfur bacterium, has two bona fide RubisCOs (form I and form II) and two RLPs (RLP1 and RLP2). Both the sequence alignment and structural analysis (discussed above) show that one of the RLPs (RLP2) is closely related to the *C. tepidum* RLP (30). Although the overall recently solved structure of *R. palustris* RLP2 is similar to that of *C. tepidum* RLP, there are subtle differences (discussed below). Disruption of either of the two RLPs present in *R. palustris* or the single RLP in *R. rubrum* does not appear to affect the expression of any oxidative stress response proteins (J. Singh, T. E. Hanson, S. Romagnoli, and F. R. Tabita, unpublished data). Because the disruption of the RLPs in these organisms failed to evoke a detectable phenotype, either these proteins do not function in the stress response or, perhaps, the amount of RubisCO present is enough to complement the function of the missing RLP (much like how RubisCO complements the defect in methionine metabolism in *B. subtilis*) (8). Interestingly, both *R. rubrum* and *R. palustris* are capable of using 5-methlythioadenosine (MTA) as the sole sulfur source for growth (Fig. 4), much like *B. subtilis*. Further studies indicate that RLP is definitely involved in MTA-dependent growth in these organisms (Singh and Tabita, unpublished). In addition, bioinformatic analyses indicate that *R. rubrum* and *R. palustris* contain the requisite genes of the methionine salvage pathway, while other nonsulfur bacteria such as *R. sphaeroides* and *R. capsulatus* do not, nor are the latter two organisms capable of MTA-dependent growth (Fig. 4). Why physiologically related organisms that are typically found in similar environments appear to both utilize (and contain RLP) and not utilize the methionine salvage pathway is not clear at this time.

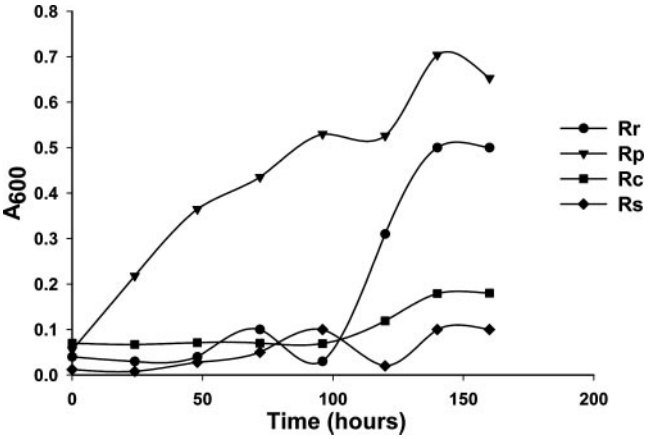


FIG. 4. Growth of four purple nonsulfur bacteria on MTA as the sole sulfur source. Rr, *Rhodospirillum rubrum*; Rp, *Rhodopseudomonas palustris*; Rc, *Rhodobacter capsulatus*; Rs, *Rhodobacter sphaeroides.* Growth on MTA correlates with the presence of RLP, further shown by inactivating the RLP gene (Singh and Tabita, unpublished).

(ii) Local gene conservation as an indicator of different functions. One method for assigning a physiological role for the functions of unassigned gene products is "guilt by association," or examining the conservation of genes that are colocalized with the gene of interest across multiple genomes. This assumes that functionally related genes will be linearly inherited or laterally transferred as conserved functional modules. This was examined for each RLP lineage and all form III RubisCOs by aligning genomes in the Integrated Microbial Genomes database against each other, centered on the gene encoding RLP (Fig. 5).

When the genomic regions surrounding the genes encoding the *C. tepidum* and *B. subtilis* RLPs were compared, distinct patterns of gene conservation were observed. In *C. tepidum* and other green sulfur bacteria, a tightly conserved core of five genes was found in seven strains, with complete sequence coverage across the area (Fig. 5B). These include two distinct short-chain dehydrogenase/reductase family homologs and two conserved hypothetical proteins, one of which displays weak similarity to predicted aldolases. A more loosely conserved, extended region upstream of the *C. tepidum* RLP-encoding gene encodes ribosomal proteins and a potential regulator (*recX*), and in three strains, there are two genes involved in bacteriochlorophyll biosynthesis. This close association with bacteriochlorophyll biosynthesis genes is intriguing, as the *C. tepidum* RLP mutant displays lowered pigment content and altered in vivo photopigment organization (31). This perturbation of photopigment organization has also been observed in

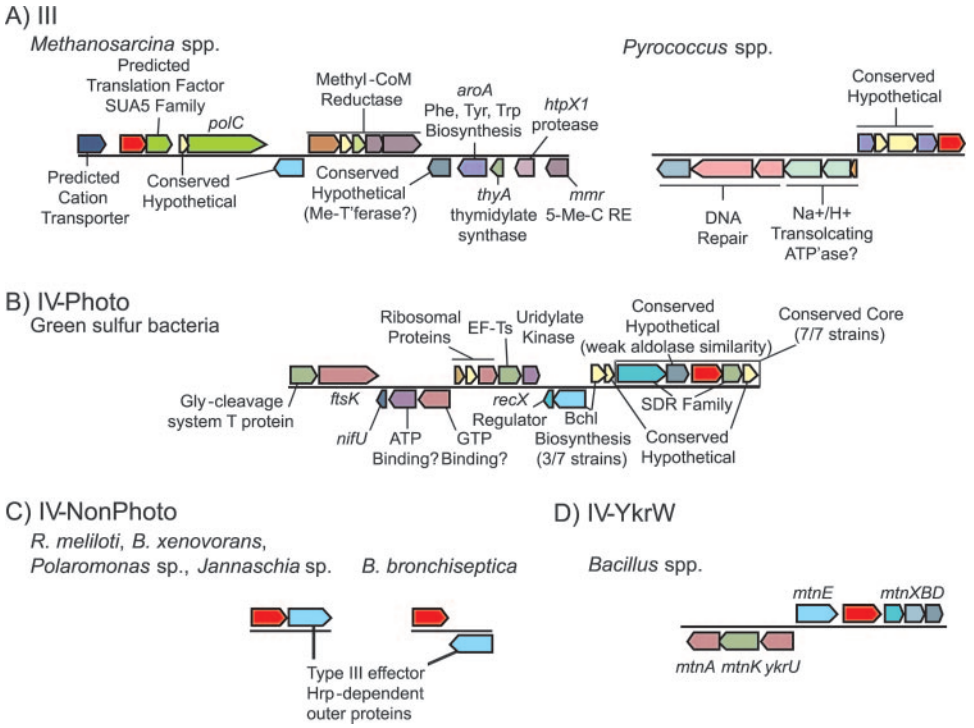


FIG. 5. Local conservation near genes encoding form III RubisCO (A) or the RLP lineages IV-Photo (B), IV-NonPhoto (C), and IV-YkrW. Gene neighborhoods were visualized using tools at the Integrated Microbial Genomes website (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). RubisCO/RLP genes are indicated in red. Other open reading frames are colored and identified according to their annotation in the Integrated Microbial Genomes database. Methyl-coM, methyl coenzyme M; Bchl, bacteriochlorophyll; Me-T'ferase, methyltransferase; 5-Me-C RE, 5-methylcytosine removing enzyme; EF-Ts, elongation factor Ts; SDR, short chain dehydrogenase/reductase.

strains of *C. tepidum* that carry mutations in potential sulfur oxidation genes other than that encoding the RLP (12). While these perturbations do not indicate a direct role for the RLP or putative sulfur oxidation genes in photopigment biosynthesis, they do indicate significant physiological shifts in the mutant strains, which can be replicated in the wild type by light and/or thermal stress (R. Morgan-Kiss and T. E. Hanson, unpublished data). In contrast, the *ykrW* gene in *B. subtilis* is surrounded by genes implicated in a pathway for the recycling of MTA, a by-product of polyamine biosynthesis (Fig. 5D). This association is consistent with the observed in vitro biochemical activity of this enzyme, e.g., enolization of the MTA salvage intermediate 2,3-diketo-5-methylthiopentyl-1-phosphate (7).

Aside from the form I and II bona fide RubisCOs and the IV-Photo and IV-YkrW lineages discussed above, only three other examples of local gene conservation were found. Form III RubisCOs in *Methanosarcina* spp. share conserved gene organizations downstream, including a methyl coenzyme (CoM) reductase operon; the *polC* gene, encoding a DNA polymerase; and others (Fig. 5A). In *Pyrococcus* spp., genes encoding form III RubisCO are preceded by four genes encoding conserved hypothetical proteins as well as potential operons encoding Na^+/H^+-translocating ATPase and potential DNA repair functions (Fig. 5A). Finally, in the IV-NonPhoto lineage, there seems to be a conserved gene encoding a surface or secreted protein predicted to be dependent on a type III secretion system for export (Fig. 5C). The functional significance of these other instances of local gene conservation is currently unknown.

Genomic Context-Based Analyses of Diverse RLPs Suggests Functional Diversity

As there are more than 300 complete genome sequences available, we used bioinformatic approaches to assist us in understanding potential functions of RLPs, most of which are uncharacterized proteins. Four genomic context-based methods were used to infer protein functions based on comparisons of hundreds of genome sequences. The phylogenetic profile method infers protein functional linkages between two proteins based on their correlated evolution in multiple genomes (53). The Rosetta Stone method infers the linkages based on the fusion of two protein-encoded genes in another genome (24, 43). The gene neighbor method assigns protein functional linkages based on the close proximity of two genes on the chromosomes in many genomes (14, 50), and the gene cluster method infers the linkages between two genes based on the operon structures in prokaryotic genomes (10, 54).

We calculated the functional linkages of 11 RLP sequences out of 44 known sequences using a confidence threshold of 0.5. Based on the functional linkages, the 11 RLP sequences can be divided into two major groups (Fig. 6). The first group consisted of the RLPs from *C. tepidum*, *R. palustris* (RLP1 and RLP2), *Archaeoglobus fulgidus*, *Mesorhizobium loti*, and *Sino-*

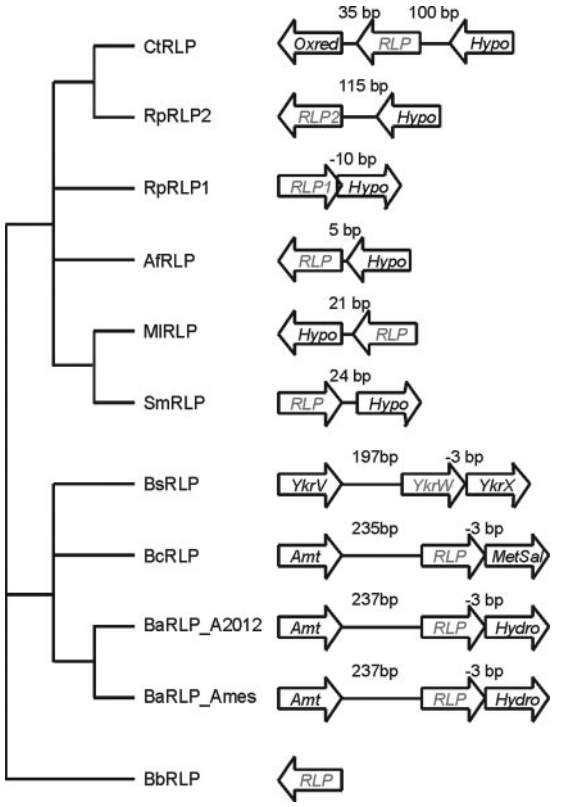


FIG. 6. RLPs grouped by their functional linkage patterns. The 11 RLPs indicated here can be divided into two major groups. In the first group, all RLPs are linked to a hypothetical protein by the gene cluster method with short intergenic distances. The two hypothetical proteins next to the RLPs in *Mesorhizobium loti* and *Sinorhizobium meliloti* are homologous to each other. All the RLPs from *Bacillus* species form the second group. They have very similar gene organizations on the chromosome. They all reside between an aminotransferase and a hydrolase, which overlaps with RLPs by 3 bp. The RLP from *Bordetella bronchiseptica* does not have any functional linkages with high confidence. Oxred, oxidoreductase; Hypo, hypothetical protein; Amt, aminotransferase; MetSal, methylthioribose salvage protein; Hydro, hydrolase, haloacid dehalogenase-like hydrolase; CtRLP, *C. tepidum* RLP; RpRLP2, *R. palustris* RLP2; AfRLP *A. fulgidus* RLP; SmRLP, *Sinorhizobium meliloti* RLP; BsRLP, *B. subtilis* RLP; BcRLP, *B. cereus* RLP; BaRLP, *B. anthracis* RLP; BbRLP, *Bordetella bronchiseptica* RLP.

rhizobium meliloti. These RLPs are all functionally linked to hypothetical proteins, which reside near the RLPs on the chromosome. The second group consisted mainly of RLPs from *Bacillus* spp., including *B. subtilis*, *B. cereus*, *B. anthracis* strain A2012, and *B. anthracis* strain Ames. These RLP genes all overlap with haloacid dehalogenase-like hydrolases with an intergenic distance of −3. In addition, they all have functional linkages to aminotransferases, which reside near the RLPs on the chromosome. Based on biochemical studies, the RLP from *B. subtilis*, YkrW/MtnW, and its two functionally linked proteins, hydrolase (YkrX/MtnX) and aminotransferase (YkrV/MtnV), have been suggested to function together in the methionine salvage pathway (8) (Fig. 7). As discussed above, YkrW/MtnW functions as an enolase with 2,3-diketo-5-methylthiopentyl-1-phosphate as its substrate. By analogy, the RLPs in the second group probably all function as enolases in the methionine salvage pathway. Certainly, the close structural likeliness of RuBP and the 2,3-diketo compound of the methionine salvage pathway (Fig. 8) and the reported ability of the *R. rubrum* RubisCO gene to complement a *ykrW* knockout in *B. subtilis* (8) suggest both a functional relationship and an evolutionary relationship between RubisCOs and RLPs. Lastly, the RLP from *Bordetella bronchiseptica* has no functional linkages above the confidence threshold and may thus belong to another group of RLPs.

In summary, relationships based on sequence similarity (see above) indicate the presence of three different lineages of bona fide RubisCO and a fourth lineage representing the RLPs that can perhaps be divided into six different subgroups. Further genetic and biochemical studies should eventually clarify the functions of each of the different RLP groups and shed further light on the evolution of RLP and RubisCO. Ultimately, the final test of functional conservation across lineages will be the heterologous expression of RLPs from different lineages in mutant strains lacking the cognate RLP for that particular organism. Early reports indicate that a form II RubisCO gene could complement a *B. subtilis* mutant lacking YkrW (8). In addition, a cyanobacterial (*M. aeruginosa*) RLP gene has also been shown to functionally complement the *B. subtilis* mutant (11). Detailed functional and structural relationships among bona fide RubisCO and RLP are extensively discussed below; clearly, bioinformatic analyses suggest discrete functions for at least some of the phylogenetically diverse RLPs discussed here.

PROBING THE EVOLUTIONARY ORIGINS OF RubisCO: EVIDENCE FOR ARCHAEAL CENTRAL METABOLISM AS THE ULTIMATE SOURCE OF ALL EXTANT RubisCO AND RLP SEQUENCES

The reconstruction of phylogenetic associations can be used to infer evolutionary relationships among related sequences. Evolutionary questions regarding the development of the bona fide RubisCOs and their relationships to RLPs are clearly of interest. The relationships between these lineages were examined with four different phylogenetic reconstruction methods (neighbor joining [NJ], minimum evolution [ME], unpaired group mean average [UPGMA], and maximum parsimony [MP]) (38) after examining amino acid distance data across all sequences via the program ProtTest to suggest an appropriate rate distribution gamma parameter (1) (Fig. 2).

In every phylogenetic reconstruction examined, the bona fide RubisCOs (forms I to III) form a coherent clade, suggesting that they share a common line of descent. With minimum evolution and neighbor joining, forms I and II are late-descending nodes in a clade where the deepest branches are form III RubisCO and two additional RubisCO sequences from *Methanosaeta thermophila* and *Methanospirillum hungatei*. These two archaeal sequences consistently clade with one another and separate from other archaeal RubisCO sequences in form III. In addition, the sequence of the RubisCO from *Methanococcoides burtonii*, a methanogenic archaeon isolated from Antarctic marine sediments (60), consistently branches at the base of the form II clade in every method employed. These sequences are quite divergent, averaging only 28% (*M. thermophila* and *M. hungatei*) and 24% (*M. burtonii*) identity with

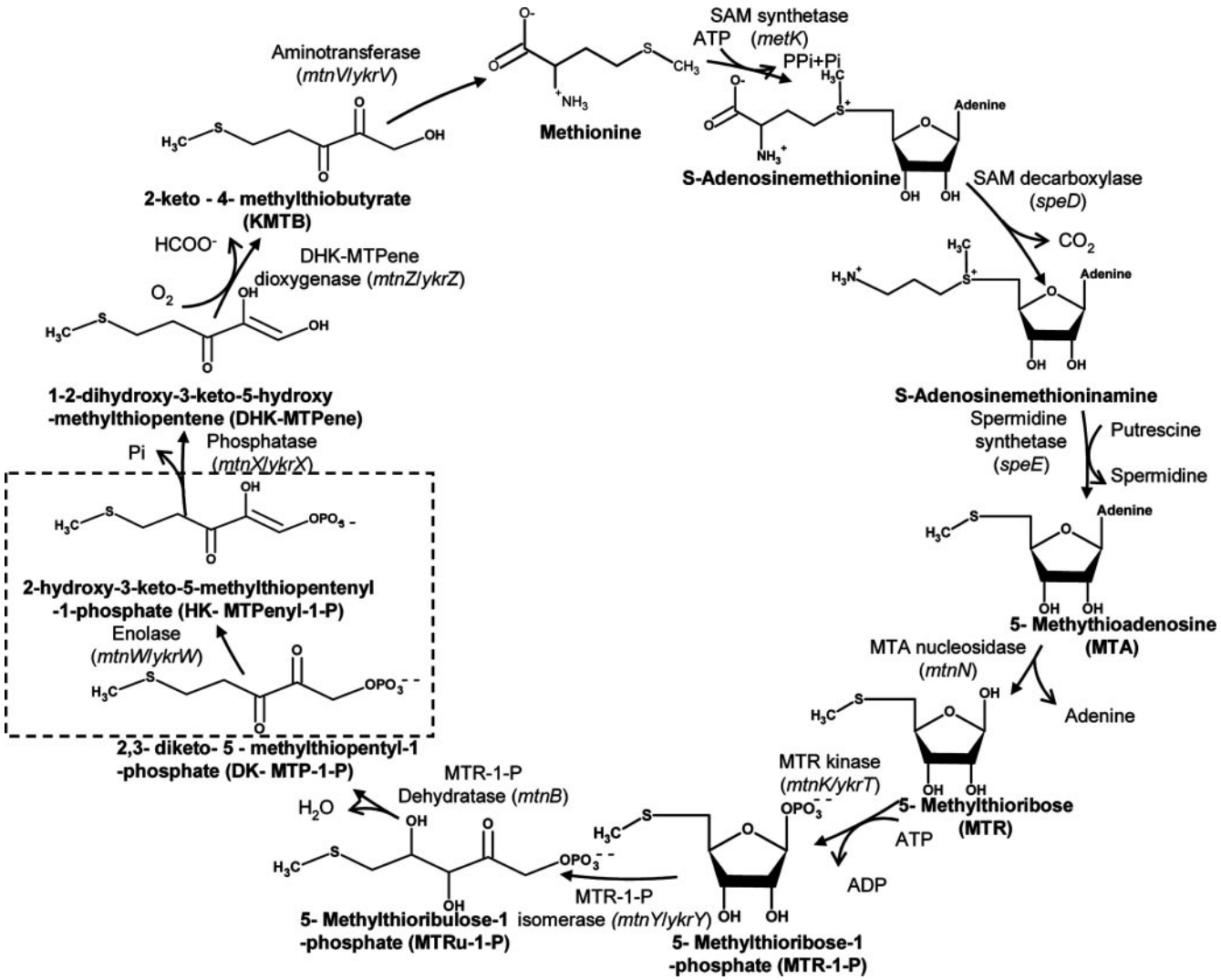


FIG. 7. Methionine salvage pathway in which the YkrW-type RLP, such as the protein from *B. subtilis* (8), encoded by the *mtnW/ykrW* gene, participates in an enolase reaction whereby 2,3-diketo-5-methylthiopentyl-1-phosphate is converted to 2-hydroxy-3-keto-5-methylthiopentenyl-1-phosphate (highlighted). The products of the *mtnX/ykrX*, *mtnZ/ykrZ*, and *mtnV/ykrV* genes then allow methionine to be formed. SAM, *S*-adenosylmethionine. (Adapted from reference 8 with permission of the publisher.)

all other RubisCO/RLP sequences. This consistent distribution of archaeal sequences at the base of clades containing all known bona fide RubisCO sequences suggests that this clade may have originated in the *Archaea* and subsequently been distributed to bacteria, eukaryotic algae, and higher plants. Overall bootstrap support is high for nodes in both methods with mean values of 75% and 83% for NJ and ME, respectively. The lowest bootstrap values were observed for internal nodes of the RLP cluster, while all terminal nodes are strongly supported.

The two other methods employed to reconstruct RubisCO/RLP relationships, UPGMA and MP, display different relationships among forms I to III. UPGMA maintains the same two lineages of form III observed by MP and NJ methods but places them as a sister group to the *M. thermophila* and *M. hungatei* sequences. This archaeal cluster is a sister clade to all form I sequences by the UPGMA method. With MP, the form III sequences are rearranged into two different lineages (III* and III** in Fig. 2D) that differ from the III-1 and III-2 lineages found by the three other methods. With this rearrangement, form I RubisCOs appear as a daughter clade nested within form III sequences. MP further produces a tree in which all clades are nested and splits the IV-Photo group found by all other methods into proteobacterial and *Chlorobium* clades. The major differences between the NJ, ME, and UPGMA trees are due to the branching order of RLP lineages, specifically whether the IV-NonPhoto lineage branches deeply off the RubisCO lineage or within the RLP lineage. Mean bootstrap support is also high for UPGMA and MP, at 80% and 100%, respectively.

Obviously, the tree topologies are highly dependent on the phylogenetic inference method employed. Both UPGMA and MP are known to be the most reliable for estimating trees in data sets where evolutionary rates are nearly constant across

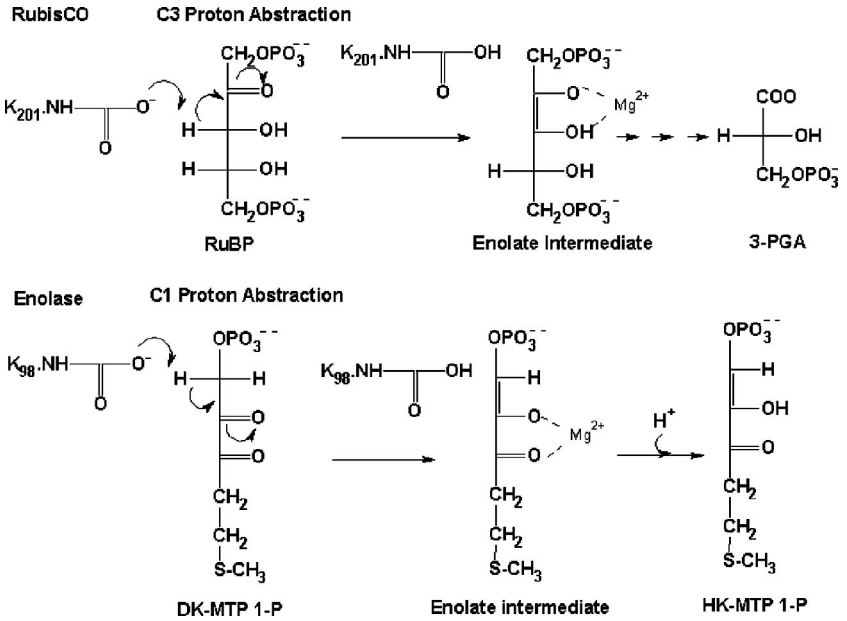


FIG. 8. RubisCO and *B. subtilis* RLP catalyze similar enolase-type reactions and employ structurally analogous substrates (see reference 33). In each instance, a carbamylated lysine catalyzes proton abstraction from the substrate to initialize enolization. DK-MTP 1-P, 2,3-diketo-5-methylthiopentyl-1-phosphate; HK-MTP 1-P, 2-hydroxy-3-keto-5-methylthiopentenyl-1-phosphate. (Adapted with permission from reference 33. Copyright 2007 American Chemical Society.)

lineages (15), while NJ with rate correction was found to operate reliably when faced with variable rates across lineages. ProtTest analysis of the RubisCO/RLP sequence set indicated a moderate amount of rate variability that could confound UPGMA and MP analyses. Thus, the phylogenetic relationships inferred by NJ and ME, which indicate an archaeal origin for RubisCO/RLP, appear to be the most robust.

RubisCO and RLP appear to be more prevalent in the euryarchaea, which, along with the crenarchaeota, are the two major branches of descent in the archaea as delineated by 16S rRNA gene sequence comparisons (18). There are only two crenarchaeal form III sequences known, those from *Hyperthermus butylicus* DSM 5456 (YP_001012710) and *Thermofilum pendens* Hrk-5 (YP_920628). Within the archaeal RubisCOs, there appears to be more flexibility in the range of residues accepted at active-site positions (Fig. 3), indicating that the final sequence of the active site is variable in this group, while it appears set in all other form I and form II sequences. Additionally, the most deeply branching RLP sequence is found in the euryarchaeon *A. fulgidus*. Taken together, these observations suggest that the euryarchaea harbor the deepest-branching RubisCO and RLP sequences, which therefore makes them the best candidates for the evolutionary root of the RubisCO and RLP superfamily.

When the phylogenetic distribution of RubisCO/RLP lineages (Table 1) was examined, a single transfer of RLP from a methanogenic euryarchaeon into an ancestor of the *Firmicutes*, *Proteobacteria*, and *Chlorobia*, with subsequent lateral transfer to chloroflexi, followed by gene losses, could account for the distribution of most of the RLP lineages. Likewise, lateral transfer of a form III RubisCO from a euryarchaeon to a common ancestor of *Cyanobacteria* and *Proteobacteria* and eukaryote RubisCOs being acquired via subsequent endosymbiotic events could account for the distribution of bona fide RubisCO lineages observed. From these considerations, the likely evolutionary development of the large subunit of RubisCO and RLP follows the model depicted in Fig. 9. In addition to this scheme, it appears that the *M. burtonii* sequence, found at the base of the form II clade, may be a result of lateral transfer of a bacterial form II sequence to the archaea.

The most recent phylogenies of the archaea based on concatenated protein trees for informational processes place the *Thermococcales* (*Pyrococcus* spp. and *Thermococcus* spp.) as the deepest-branching euryarchaeal group. Within the form III sequences, the *Thermococcales* sequences form a coherent clade with sequences from the *Haloarchaea* and *A. fulgidus* (III-1) that is separate from sequences in the methanogenic euryarchaea (III-2) (Fig. 9). This suggests that form III RubisCO may have arisen concomitantly with the divergence of the euryarchaea. A sequence from the genome of the methanogenic euryarchaeon *Methanoculleus marisnigri* falls within clade III-1, and three other methanogenic euryarchaea harbor RubisCO sequences that cannot be concretely assigned to a clade, suggesting that the methanogenic archaeal sequences are near the root of the RubisCO large-subunit superfamily. Furthermore, the only archaeal RLP found thus far is in *A. fulgidus*, a later-branching euryarchaeon that also encodes a form III RubisCO. If RubisCO/RLP evolution parallels the evolution of archaea in general, it would suggest that a form III RubisCO arising within the *Methanomicrobia* was the ultimate source of all RubisCO and RLP lineages (Fig. 9).

The scenario outlined above and in Fig. 9 does not explain the presence of RLPs in the picoeukaryote marine chlorophyte *O. tauri*, which encodes two distinct and highly divergent RLPs in its nuclear genome (17) in addition to a typical form I large

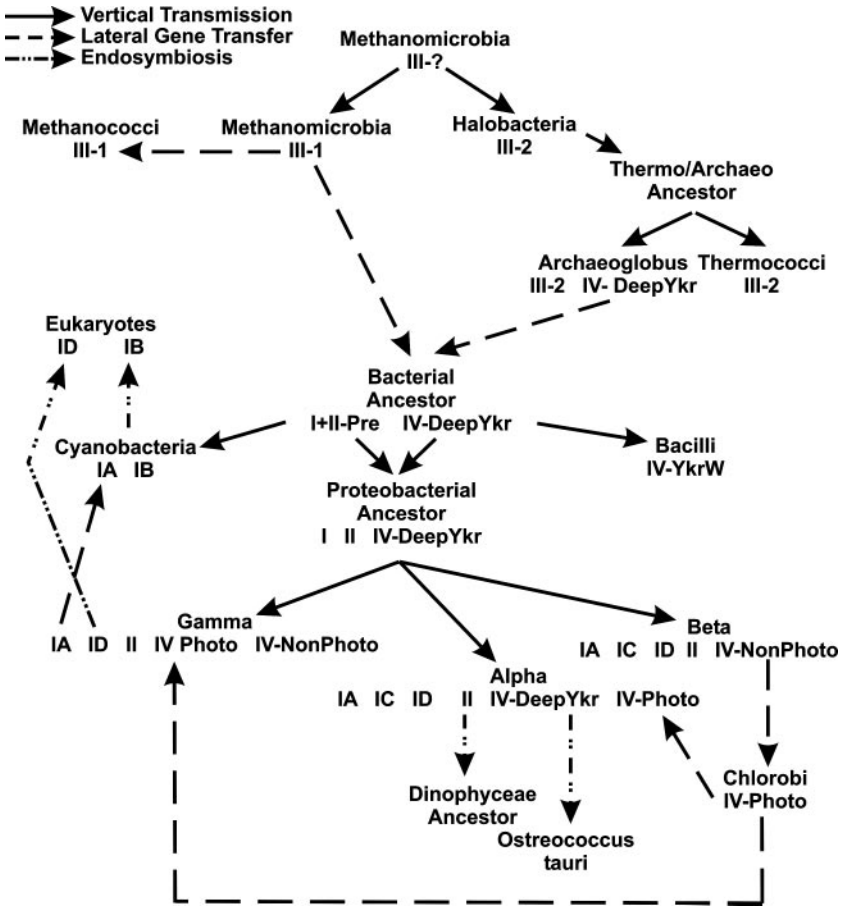


FIG. 9. Model for the evolution of RubisCO large subunits and RLP. The ancestor of all extant RubisCO large subunits and RLPs is proposed to have arisen in the *Methanomicrobia* with subsequent distribution by vertical transmission (solid arrows) and lateral transfer (dashed arrows) within the archaea. A central event in the evolutionary history was the acquisition of both a form III RubisCO and an RLP (IV-DeepYkr) by an ancestral eubacterium from the archaea. From these two ancestral sequences, diverse form I, form II, and form IV enzymes evolved within the *Proteobacteria* and *Cyanobacteria* and have been subsequently distributed by lateral gene transfer and by endosymbiotic events (dashed and dotted arrows) involving both *Cyanobacteria* and *Alphaproteobacteria*, leading to the phylogenetic distribution of sequences seen in nature today. The small subunit of form I RubisCO must have originated soon after the transfer of form III to the eubacterial ancestor prior to the divergence of *Proteobacteria* and *Cyanobacteria*.

subunit in the chloroplast genome (57). The synthesis and function of the RLPs encoded by the *O. tauri* nuclear genome have not yet been demonstrated. It may be that this sole example of eukaryotic RLPs is indicative of an additional lateral gene transfer, possibly from a member of the *Alphaproteobacteria*. The other unusual prokaryote-to-eukaryote lateral transfer in this scheme, which explains the presence of form II RubisCO in the *Dinophyceae*, has been previously analyzed in detail (44, 51, 52, 58).

The archaeal origin model of RubisCO/RLP evolution proposed here is substantially different from that reported previously by Ashida et al. (7) and Carre-Mlouka et al. (11), who speculated that bona fide RubisCOs arose in the YkrW lineage. However, those authors relied on much smaller sets of sequences and more limited numbers of phylogenetic reconstructions to reach their conclusions. Clearly, models of RubisCO evolution will themselves need to evolve, as new sequences are continuously being reported, especially in metagenomic sequencing projects. However, at this point, no new distinct RubisCO forms have been uncovered; thus, the basic conclusions reached in Fig. 9 appear to represent the most feasible scenarios for RubisCO and RLP evolution.

Non-RubisCO/RLP Structural Homologs

In attempts to reconstruct the evolution of the RubisCO/RLP superfamily, identification of a non-RubisCO/RLP-related sequence that could serve as a root or source of the superfamily has thus far been problematic. Even using score filtering to eliminate highly similar sequences from traditional similarity searches like BLAST, PHI-BLAST, and PSI-BLAST, the number of RubisCO/RLP sequences currently in databases makes for a complicated morass of sequence data that must be navigated to identify potential non-RubisCO/RLP sequences related to any query. By contrast, structural homology searches may be a useful method for identifying distantly related proteins since the degree of sequence similarity of conserved protein structural elements may be too low to be detected by typical algorithms. Moreover, the search space is also much less crowded, as relatively few structures have been determined

TABLE 3. Structural homologs of RubisCO/RLP as determined by fold comparisons carried out at the DALI server with five representative RubisCO and RLP structures against the PDB database[a]

PDB accession no.	Mean *Z* score	SCOP superfamily(ies)
1GK8	44.6	RubisCO, C-terminal domain, and large subunit, small (N-terminal) domain
1TEL	49.5	RubisCO, C-terminal domain, and large subunit, small (N-terminal) domain
2RUS	41.9	RubisCO, C-terminal domain, and large subunit, small (N-terminal) domain
1KV8	15.0	Ribulose phosphate binding barrel
2TYS	13.9	Ribulose phosphate binding barrel and tryptophan synthase beta subunit-like PLP-dependent enzymes
1RPX	13.9	Ribulose phosphate binding barrel
1NAL	13.6	Aldolase
1QFE	13.0	Aldolase
1EZW	12.3	Bacterial luciferase-like
1TVL	12.1	Bacterial luciferase-like
1M41	11.5	Bacterial luciferase-like
1T8Q	10.6	PLC-like phosphodiesterase

[a] PDB accession numbers 1RBL (form I), 5RUB (form II), 1GEH (form III), 1YKW (form IV/RLP), and 2OEJ.

compared to the numbers of gene sequences available. Structurally, all RubisCO and RLP structures solved to date are members of the triose phosphate isomerase (TIM)/mutase fold, characterized by an eight-membered β/α-barrel motif, the TIM barrel (46, 76). The TIM barrel fold is composed of 32 superfamilies in the latest release of the Structural Classification of Proteins (SCOP) database (version 1.71 [http://scop.mrc-lmb.cam.ac.uk/scop/]), the largest number of structural superfamilies of any fold within the SCOP class of alpha and beta proteins (α/β). The functional flexibility of the TIM barrel scaffold has been well documented (reviewed by Anantharaman et al. [3]). The evolution of TIM barrel proteins has been previously examined, and the RubisCO superfamily was found to cluster with other TIM barrel superfamilies containing a sugar phosphate binding motif (76). However, a separate PSI-BLAST analysis failed to link the RubisCO structure with other TIM families (46).

To identify structural homologs of RubisCO and RLP, a total of five RubisCO/RLP structures representing each major lineage (Protein Data Bank [PDB] accession numbers 1RBL [form I], 5RUB [form II], 1GEH [form III], 1YKW [form IV/RLP], and 2OEJ) were used to search the PDB structure collection using the DALI fold comparison search tool (http://www.ebi.ac.uk/dali/index.html). The DALI server was chosen based on its favorable evaluation relative to other fold comparison servers (48). Structural homologs were considered only if the average DALI *Z* score was >10 and the structure was identified by each of the five queries (Table 3). The number of homologs returned was driven primarily by PDB accession number 5RUB, the form II RubisCO structure from *Rhodospirillum rubrum*, which retrieved the fewest homologs.

As expected, additional RubisCO structures were identified as the closest structural homologs of the queries, with an average *Z* score of 45.3 for the structures under accession numbers 1GK8 (form I RubisCO from *Chlamydomonas reinhardtii*) (72), 1TEL (an independently solved structure of *C. tepidum* RLP) (33), and 2RUS (activated complex of the *R. rubrum* form II enzyme) (42). Beyond other RubisCO and RLP structures, the detected structural homologs were all superfamilies within the TIM barrel fold. The three closest homologs were from the ribulose phosphate binding barrel superfamily, PDB accession numbers 1KV8 (3-keto-L-gulonate-6-phosphate decarboxylase from *E. coli*) (77), 2TYS (tryptophan synthase from *Salmonella enterica* serovar Typhimurium) (56), and 1RPX (D-ribulose-5-phosphate-3-epimerase from *Solanum tuberosum* chloroplasts) (36). Two of these enzymes play key roles in central metabolic pathways of amino acid biosynthesis or sugar phosphate interconversions of the pentose phosphate pathway, as do most other members of this SCOP superfamily (see http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.d.b.c.html for details) (6). As expected for central metabolic enzymes, homologs of these sequences are encoded by nearly all archaeal genomes (see http://img.jgi.doe.gov/cgi-bin/pub/main.cgi for details). As outlined below, the form III RubisCO that we propose to be the evolutionary source of all other RubisCO and RLP sequences appears to have evolved to reclaim potentially dead-end five-carbon sugar bisphosphates and salvage them to central metabolic pathways. This structural comparison suggests that a second phosphate binding site may have been the key step in the evolution of RubisCO and that the source protein for RubisCO could have been recruited from a common, central metabolic pathway. Newer statistical methods of long-range phylogenetic reconstruction (i.e., hidden Markov models) may provide support for the structural comparison arguments posed above and identify specific candidates as the ultimate sources for the RubisCO superfamily.

Physiological Role for Archaeal (Form III) RubisCO

Previous studies have shown that genes that encode catalytically competent recombinant RubisCO (form III) are present in some archaea (25, 73), and the protein appears to be functional in some organisms (27). However, the lack of any demonstrable phosphoribulokinase (PRK) activity (or a gene that encodes this protein) from these same organisms that contain RubisCO has been a major curiosity, as such organisms seemingly would not possess a means to synthesize the unique keto sugar (RuBP) that is the substrate for RubisCO. This conundrum was recently addressed, with two possibilities considered: (i) form III archaeal RubisCO preferentially uses an alternative substrate and does not require RuBP for catalysis, or (ii) alternative means to synthesize RuBP that are unique to archaea exist. The first possibility, if true, would also suggest that RuBP-dependent RubisCO activity might have evolved from a protein that possessed some alternative activity, a theory espoused by those that believe that RLP is an evolutionary precursor to RubisCO (7, 11). However, exhaustive studies have thus far found no alternative to RuBP as a substrate or CO_2 acceptor for archaeal (or any other) RubisCO (26), suggesting that it is unlikely that form III RubisCO is being used for anything other than producing 3-phosphoglyceric acid (PGA) from RuBP and CO_2. In addition, the mesophilic archaeal RubisCO gene complemented a RubisCO deletion mutant of *Rhodobacter*

capsulatus to autotrophic growth, showing that RuBP is also a substrate for this enzyme in vivo (27).

With evidence pointing to RuBP as the exclusive substrate for archaeal RubisCO, how, then, does the organism synthesize RuBP? Again, negative data suggest that there is no demonstrable PRK activity in extracts from the organisms tested (26); moreover, analyses of the great majority of available genomes indicate no recognizable gene to encode PRK. Recently determined genomic sequences of *Methanospirillum hungatei*, *Methanoculleus marisnigri*, and *Methanosaeta thermophila* (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) represent the only archaeal organisms where potential PRK genes may exist. As for the vast majority of archaea, which possess no discernible PRK gene, a satisfying positive finding was the demonstration of a novel means to synthesize RuBP. Direct enzymatic assays using alternative substrates with extracts of *Methanocaldococcus jannaschii* provided evidence for a previously uncharacterized pathway for RuBP synthesis from 5-phosphoribose-D-1-pyrophosphate (PRPP) in *M. jannaschii* and other methanogenic archaea (26). Thus, these experiments, using PRPP as the sole substrate, resolved the need for a kinase dedicated to RuBP generation because PRPP already contains the relevant phosphates at both the C1 and C5 positions. Based on studies with other systems, it was hypothesized that either there is a selective enzymatic dephosphorylation step at the C1 position or nonenzymatic dephosphorylation occurs at the pyrophosphate at both moderate and high temperatures in the presence of magnesium at neutral pH (26). In either instance, the product would be ribose-1,5-bisphophosphate, a compound known to be synthesized in many other biological systems including macrophages and red blood cells under conditions of hypoxia (discussed in reference 24). Further indications of a novel and specific enzymatic reaction(s) was the stoichiometric conversion of PRPP to RuBP using extracts of *M. jannaschii*, such that that one molecule of PRPP was converted to two molecules of PGA. These results provided experimental verification for the proposed pathway (26). Inhibition of the PRPP-to-PGA conversion in vitro by both the RubisCO transition state analog CABP and antibodies to *M. jannaschii* RubisCO convincingly reinforced the idea that RubisCO catalysis is essential to convert PRPP to PGA. The proposed unique enzymatic step of this pathway is the conversion of ribose-1,5-bisphosphate, or ribose-1,2 cyclic phosphate-5-P (ribose-1,2cP-5-P), to RuBP. This work thus identified a novel means to synthesize the CO_2 acceptor and substrate for RubisCO in the absence of a detectable kinase such as PRK. More recently, studies with *Thermococcus kodakarensis* confirmed and greatly extended those studies and again pointed to ribose-1,5-bisphosphate as the direct precursor to RuBP (59). Moreover, Sato et al. identified the enzymes and the requisite structural genes, including RubisCO, that are involved in a pathway of AMP metabolism (59). In that scheme, AMP, which could be produced from PRPP, is acted upon by an AMP phosphorylase to produce ribose-1,5-bisphosphate, followed by a ribose-1,5-bisphosphate isomerase, to yield RuBP. Both studies proposed that this route to RuBP might point to unique evolutionary links between purine-pyrimidine recycling pathways and the CBB cycle, with RubisCO catalysis and PRPP/AMP metabolism providing the needed anaplerotic levels of PGA (26, 59). Apparently, the genes of the AMP-to-RuBP pathway are conserved in virtually all archaea that contain form III RubisCO (59), suggesting that this might be a universal means by which archaea employ RubisCO in metabolism.

RubisCO AND RLP STRUCTURES: SIMILAR YET DIFFERENT ENOUGH

The structure of the RLP from *C. tepidum* was recently solved to a resolution of 2.0 Å and shown to be a homodimer of large subunits, similar to various form II and form III RubisCOs (39). Indeed, the overall secondary structures of individual monomeric units of bona fide RubisCOs from all sources and form IV (RLP) are quite similar. Each subunit of *C. tepidum* RLP is composed of a smaller N-terminal domain and a larger C-terminal domain. The N-terminal domain, residues 1 to 45, consists of a four-stranded β-sheet with helices on one side of the sheet. The C-terminal domain, residues 146 to 435, consists of an eight-stranded α/β-barrel with two additional small α-helices forming a cap at the C terminus (39). Like form I, form II, and form III RubisCOs, the presumptive active site of RLP is located in the subunit interface between the C-terminal domain of one subunit and the N-terminal domain of another subunit (Fig. 10). As discussed above, compared to the invariant active-site residues found in RubisCO, 10 of 19 active-site residues differ in the *C. tepidum* RLP. These dissimilarities in the amino acid sequence confer unique shapes and chemical properties to the active site, making it evident that *C. tepidum* RLP may not bind RuBP but may bind a structurally related molecule. While the *C. tepidum* RLP active site appears to be compatible for accommodating the P1 phosphate group, the backbone of CABP, and a metal ion (possibly Mg^{2+}), the geometry and chemistry seem to be incompatible with an incoming P2 phosphate group, as in CABP (39). It does appear, however, that a smaller and slightly hydrophobic group may fit into this active site. One is also confronted with the interesting result that *R. rubrum* RubisCO complements the function of *B. subtilis* RLP (YkrW/MtnW) in a *ykrW/mtnW* mutant (8). Since the substrates for the 2,3-diketo-5-methylthiopentyl-1-phosphate enolase and RubisCO reactions are fairly similar (Fig. 8), one would expect that the active sites of RLP should be able to bind to a wide range of molecules similar to RuBP, as is the case with RubisCO (5). However, structural analyses of *C. tepidum* RLP indicate that nonidentical residues at positions coincident with mechanistically significant RubisCO residues make the RLP active-site pocket smaller and slightly more hydrophobic at the P2 site (39) (Fig. 10) and hence may cause steric hindrance for binding the P2 phosphate of RuBP. The active-site structure of *C. tepidum* RLP suggests that this protein might function as an enolase but probably could not catalyze carboxylation (39).

Loop 6, which is in the C-terminal α/β-barrel domain of RubisCO, plays an important role in catalysis (13). Among multiple form I RubisCO structures, loop 6 has been observed to partition between the "open" and "closed" conformations (20, 61, 62). In *C. tepidum* RLP, loop 6 is ordered and adopts a closed conformation similar to that found in the structure of activated RubisCO (PDB accession number 8RUC), although no substrate is bound at the active site. Loop 6 folds over and

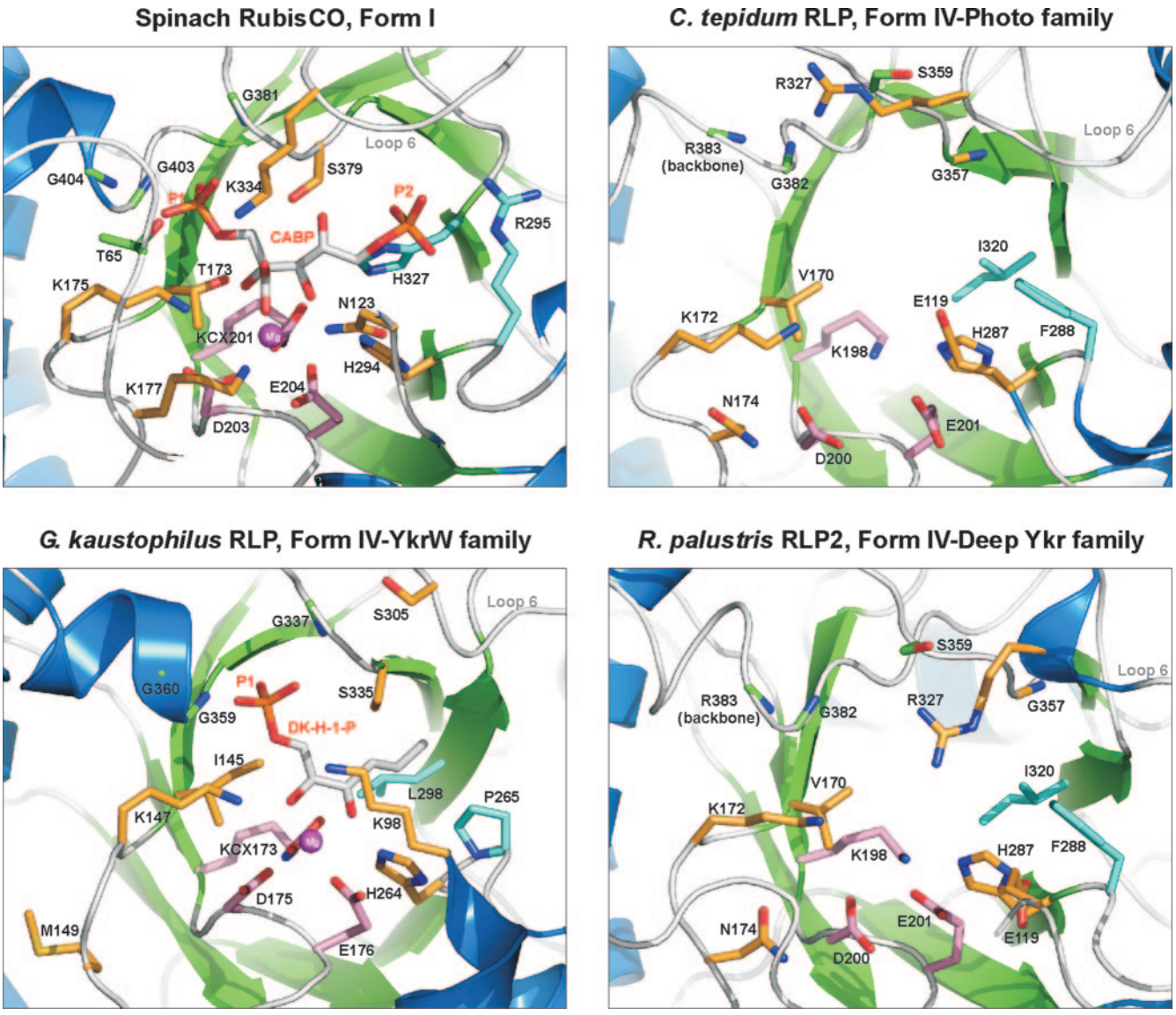


FIG. 10. The active sites in the crystal structures of form I (spinach) RubisCO (PDB accession number 8RUC) with bound CABP, *C. tepidum* RLP (PDB accession number 1YKW), *R. palustris* RLP2 (PDB accession number 2QYG), and *G. kaustophilus* RLP (accession number 2OEM) with bound DK-H-1-P. The side chains of active-site residues are shown as sticks, except for residue R383 in *C. tepidum* RLP and *R. palustris* RLP2. Only the backbone carbon and nitrogen atoms of R383 in the RLPs are shown. CABP and DK-H-1-P are shown in white, and the P1 and P2 phosphate groups are labeled in red and orange. Residues involved in contributing hydrogen bonds with the P1 phosphate group are green, residues involved in making hydrogen bonds with the backbone of CABP are orange, residues coordinating the Mg^{2+} atom (shown in magenta) are light red, and residues involved in binding P2 phosphate group are cyan. Not all parts of the structures are shown for the purpose of clarity.

closes the active site. The backbone of a key residue on loop 6, Arg-327, superimposes well with that of Lys-334 in form I RubisCO (Fig. 10). Although the side chain has a different conformation, it can possibly make hydrogen bonds with CABP. There are two other major differences in the *C. tepidum* RLP structure compared to those of bona fide RubisCO proteins. First, there is an additional 14-residue loop, loop CD, between β-strands C and D in the N-terminal domain. Second, *C. tepidum* RLP is missing a β-hairpin turn between helix 6 and β-strand 7 in the C-terminal α/β-barrel domain. Loop CD approaches the active-site opening from the direction opposite from loop 6 and packs against loop 6. The positional similarity between loop CD and the C-terminal tail of RubisCO upon substrate binding suggests that loop CD may have a role in positioning loop 6 (discussed below).

In this review, we also report the second crystal structure of an RLP from the IV-Photo clade, RLP2 from *R. palustris* (PDB accession number 2QYG) (Table 4). As described above, the structure of *R. palustris* RLP2 is very similar to the structure of *C. tepidum* RLP, with a Cα atom rmsd of 0.8 Å (Fig. 11). A previously described disordered region in N-terminal domain residues 47 to 58 was found to be ordered in the *R. palustris* RLP2 structure, as in another independently solved *C. tepidum* RLP structure (cited in reference 31). In addition, much like *C. tepidum* RLP, the same residues analogous to RubisCO active-site residues are conserved in *R. palustris* RLP2. However, in *R.*

TABLE 4. X-ray data collection and refinement statistics of the *R. palustris* RLP2 structure

Parameter[b]	Value
Wavelength (Å)	0.9537
Temp (K)	100
Space group	$P2_12_12_1$
Cell parameter (Å)	
a	68.66
b	119.53
c	203.04
Resolution (Å)	103.1–3.3 (3.4–3.3)
No. of reflections	
Total	163,935
Unique	26,399
Completeness[a] (%)	99.7 (99.5)
$<I/\sigma>$[a]	9.6 (3.8)
R_{sym}[a] (%)	16.4 (46.7)
Model refinement	
R/R_{free}[a] (%)	20.3/23.2 (26.8/30.1)
No. of protein atoms	13,016
RMS bond length (Å)	0.012
RMS bond angles (°)	1.396
Ramachandran plot [no. of residues (%)]	
Most favored	1,236 (88.0)
Additional allowed	147 (10.5)
Generously allowed	17 (1.2)
Disallowed	4 (0.3)

[a] Statistics for the outer resolution shell are given in parentheses.
[b] $R_{sym} = \Sigma(I - <I>)^2/\Sigma I^2$; *R*, *R* factor; R_{free}, subset of reflections not included in structure refinement; *I*, intensity of reflections; RMS, root mean square.

palustris RLP2, residue R327 appears to take up a different conformation compared to that in the *C. tepidum* RLP structure (Fig. 10). In *R. palustris* RLP2, the side chain of R327 adopts a conformation comparable to that of K334, the corresponding catalytic residue of spinach (form I) RubisCO (PDB accession number 8RUC). The *R. palustris* RLP2 structure further supports the hypothesis that residue R327 can potentially form hydrogen bonds with the P1 phosphate and the backbone of a CABP-like ligand. Residue E119, although adopting a different conformation relative to the identical residue in *C. tepidum* RLP, can still potentially form a hydrogen bond with the backbone of a CABP-like ligand.

More recently, structure-function studies of a YkrW-type RLP from *Geobacillus kaustophilus* (previously *Bacillus stearothermophilus*) established the structural basis for the "enolase" function of YkrW. Evidence points to the involvement of Lys-98 in proton abstraction, with this residue likely serving as the general base during catalysis, much as Lys-201 (or its equivalent in different forms) serves as the general base during RubisCO catalysis (Fig. 8) (33). Interestingly, Lys-173 of *B. kaustophilus* RLP, which is structurally analogous to Lys-201, is also carboxylated and coordinates with Mg^{2+} when a substrate analog is bound. The compound 2,3-diketohexane-1-phosphate (DK-H-1-P), a substrate analog of 2,3-diketo-5-methylthiopentyl-1-phosphate, was shown to be bound to *G. kaustophilus* RLP in a manner similar to that of the binding of 2-CABP to RubisCO's active site, providing further credence to the conserved means by which the YkrW "enolase" and RubisCO initiate catalysis albeit with differently positioned Lys residues serving as the general bases. While Lys-98 is highly conserved in the YkrW group of proteins, three of which have now been shown to act as enolases in the methionine salvage pathway (7, 11, 33), it is clear that other RLPs possess different residues in this position, especially asparagine, identical to the conserved Asn-123 in bona fide RubisCOs (Fig. 3). This may be a further indication of different functions for RLP in organisms that lack a methionine salvage pathway. Alternatively, for organisms that do utilize RLP in a presumptive methionine salvage pathway, such as *R. rubrum* and *R. palustris* (Fig. 4), but that do not possess a Lys in this position (Fig. 3), it will be interesting to determine the general bases and their locations. In *Bacillus clausii*, an apparent YkrW-type RLP substitutes an arginine for the Lys in position 98 (Fig. 3), raising the question of whether this protein is active in the enolase reaction or perhaps uses Arg as the base to initiate the reaction or even catalyzes some alternative reaction. Likewise, the Glu in position 119 of the *C. tepidum* protein is intriguing, and, from structural considerations, it was suggested that this protein could utilize some unknown ketose phosphate substrate (33), much like the above-described analyses that indicated that this protein binds a substrate that is similar to yet smaller than RuBP (39). The residue dissimilarities at the P2 binding site in

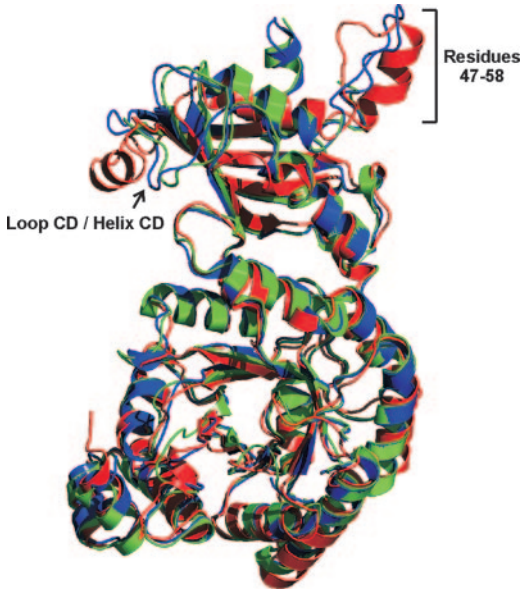


FIG. 11. The monomer structures of RLP2 from *R. palustris* and RLP from *G. kaustophilus* superimposed with the RLP from *C. tepidum*. *R. palustris* RLP2 is blue, *G. kaustophilus* RLP is red, and *C. tepidum* RLP is green. The root mean square deviation (RMSD) of the Cα atom is 0.8 Å between *R. palustris* RLP2 and *C. tepidum* RLP, 1.3 Å between *R. palustris* RLP2 and *G. kaustophilus* RLP, and 1.3 Å between *C. tepidum* RLP2 and *G. kaustophilus* RLP. Two main structural differences can be seen in the N-terminal domain: loop CD in *C. tepidum* RLP and *R. palustris* RLP2 becomes a helix in *G. kaustophilus* RLP, and residues 47 to 58, missed in *C. tepidum* RLP, become a loop in *R. palustris* RLP2 and partly a helix in *G. kaustophilus* RLP.

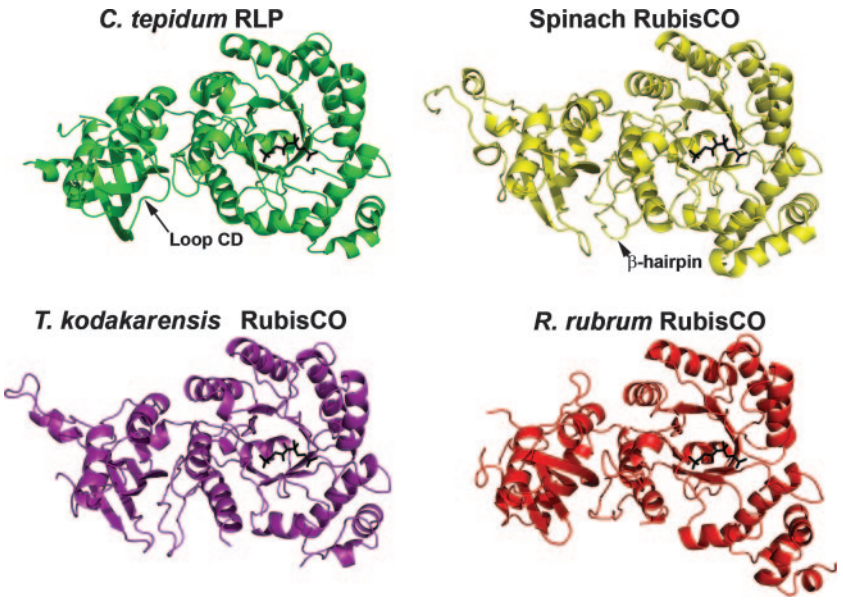


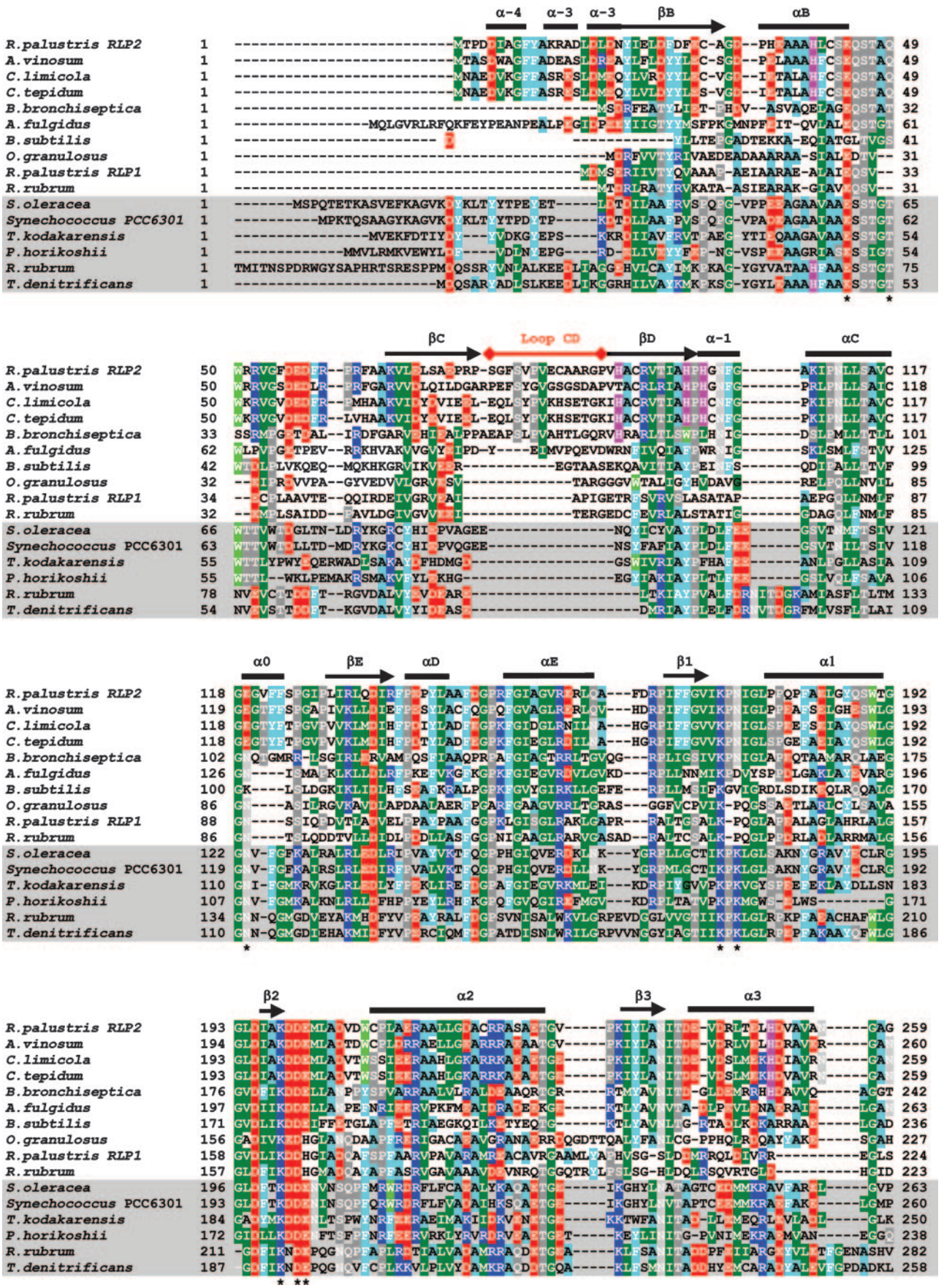
FIG. 12. Comparison of secondary structural elements in the X-ray crystal structures of the different forms of RubisCO. Large subunits from the structures of spinach (form I; PDB accession number 8RUC) (yellow), *T. kodakarensis* (form III; accession number 1GEH) (purple), and *R. rubrum* (form II; accession number 5RUB) (red) were superimposed on *C. tepidum* RLP (form IV; accession number 1YKW) (green) to align the α-carbon backbones. The transition state analog CABP (black sticks), which is present only in the spinach structure, has been drawn into the other structures to indicate the positions of active sites. A basic unit common to all four types of structures is formed as a result of the association of at least two of the large subunits. The active sites in bona fide RubisCO enzymes are contributed by residues from the N-terminal domain of one large subunit and the C-terminal domain of the other. Loop CD, which is present only in the RLPs and the RubisCO β-hairpin structure that is absent in the RLP structure, is indicated.

G. kaustophilus RLP, proline versus phenylalanine and leucine versus isoleucine, compared to the Photo-type RLP, suggest that the substrates for the RLPs from YkrW and Photo families may be similar at the P1 site and in the backbone but differ at the P2 site with different hydrophobicities and sizes (Fig. 10).

The N-terminal 18 residues in the Photo-type RLPs are missing in the YkrW type. In addition, there are two main differences between the structures of the Photo-type RLP and YkrW-type RLP in the N-terminal domain. Loop CD in *G. kaustophilus* RLP becomes a helix and slightly swings away from loop 6 but forms a tighter interaction interface with loop CD (which should be helix CD in this case) from the other monomer. The second main difference is in the region of residues 47 to 58, which was previously missing from the *C. tepidum* RLP structure. This region is less flexible in the structure of *R. palustris* RLP2, forming a loop, and partly becomes a helix in the structure of *G. kaustophilus* RLP (Fig. 11). Loop 6 in *G. kaustophilus* RLP adopts a closed conformation, as seen in the Photo-type RLPs, when the substrate analog or Mg^{2+} is bound but becomes flexible when no substrate or Mg^{2+} is bound; in addition, the density for residues 304 to 308 is missing in the latter *G. kaustophilus* structure, presumably because of the flexibility of loop 6 (PDB accession number 2OEJ).

Potential of Structural Comparisons To Enhance Functional Studies

Combinations of several techniques such as structural analysis, sequence alignments, site-directed mutagenesis, and chemical modifications have been used to deduce the roles of several active-site residues in RubisCO (reviewed in references 5 and 66). Although we lack knowledge regarding the functions of different RLPs, the overall structural similarity and yet the subtle differences that they share with bona fide RubisCO enzymes (discussed above), present us with a strong rationale for carrying out genetic engineering studies to facilitate an understanding of the physiological roles of RLPs. Since structurally divergent RubisCO enzymes all employ a common reaction mechanism, it is not unreasonable to expect that subtle changes in the RLP structure could perhaps introduce catalytic competency for RuBP carboxylation/oxygenation. The most obvious targets for genetically engineering such changes are those RubisCO active-site residues that have been altered in the RLPs. Although nonidentical, only a few of these residue identities are nonconservative relative to the nature of the side chains of their RubisCO counterparts. Taking the *C. tepidum* RLP as an example, the nonidentical active-site residues are Q49, E119, N174, F288, I320, R327, G357, S359, and R383. The corresponding residues are T65, N123, K177, R295, H327, K334, S379, G381, and G404, respectively, in the form I (spinach) RubisCO (Fig. 10). In this case, the F288R and I320H substitutions, which are the residues that bind the P2 phosphate group of RuBP or CABP in RubisCO, will be expected to have a marked effect on the structure of the active site as well as the chemical nature of the active-site pocket. The finding that F288 and I320 of *C. tepidum* RLP cannot support P2 phosphate binding is experimentally supported by the failure to detect any RubisCO activity or CABP binding to this protein (31). Certainly, substitutions at any of these conserved residues in bona fide RubisCO enzymes have always been found to have


α-4 α-3 α-3 βB αB
βC Loop CD βD α-1 αC
α0 βE αD αE β1 α1
β2 α2 β3 α3
R.palustris RLP2
A.vinosum
C.limicola
C.tepidum
B.bronchiseptica
A.fulgidus
B.subtilis
O.granulosus
R.palustris RLP1
R.rubrum
S.oleracea
Synechococcus PCC6301
T.kodakarensis
P.horikoshii
R.rubrum
T.denitrificans

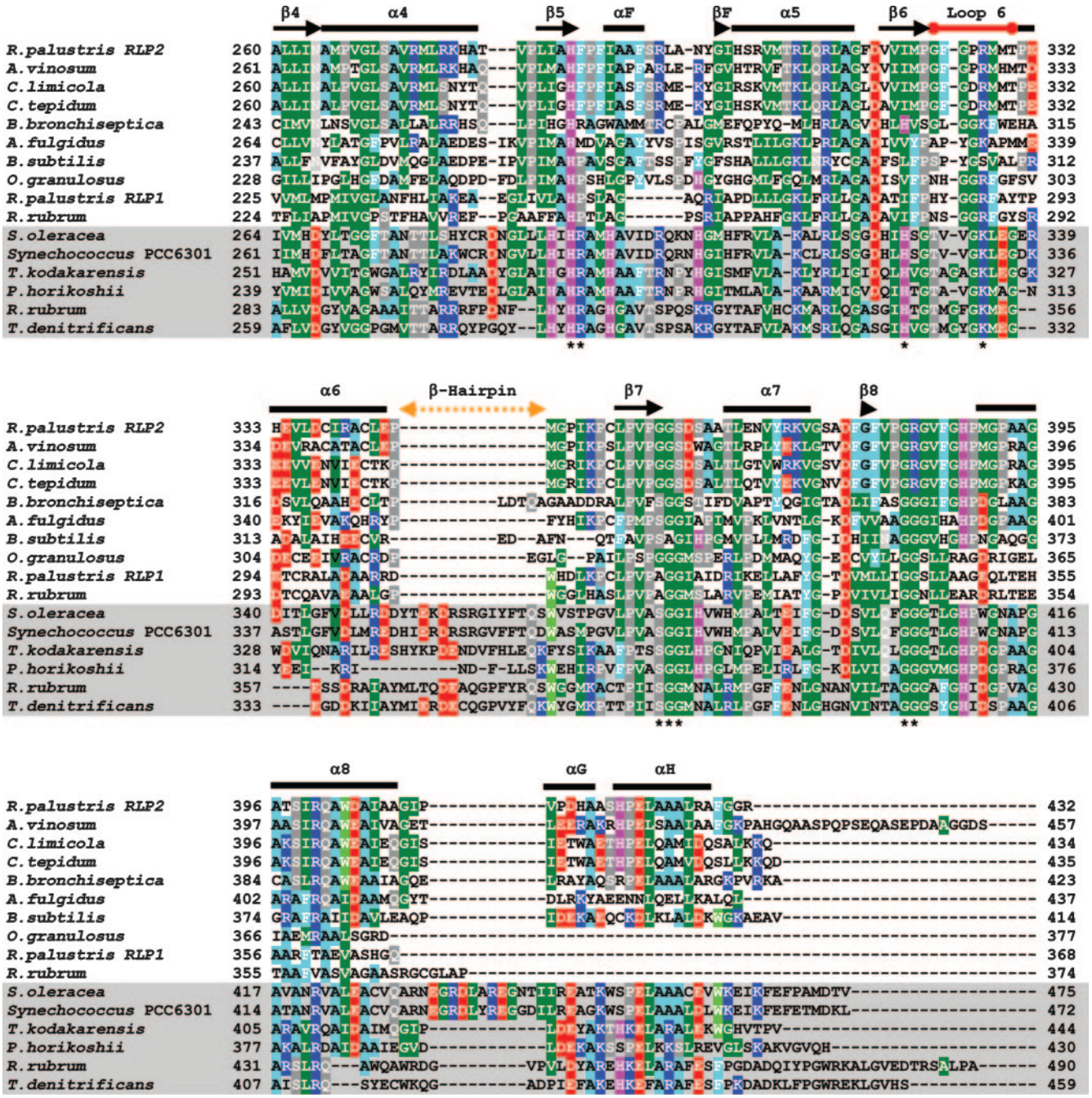


FIG. 13. Structural alignment of representative sequences from RLPs and RubisCO large subunits. Superimposition of the X-ray crystal structures of *C. tepidum* RLP (PDB accession number 1YKW; form IV), spinach RubisCO (accession number 8RUC; form I), *T. kodakarensis* RubisCO (accession number 1GEH; form III), and *R. rubrum* RubisCO (accession number 5RUB; form II) was used to deduce the alignment of secondary structural elements (helices as bars and β-strands as arrows). Residue numbers are indicated on each side of the sequences. Conserved active-site residues are marked with an "*" below the sequences. RubisCO large-subunit sequences are boxed in gray. Residues that are identical or similar to those in other species are colored uniquely based on the nature of the residue. The catalytic loop 6, β-hairpin (both present in RubisCO enzymes), and loop CD (present only in RLPs) are indicated. *A. vinosum*, *Allochromatium vinosum*; *C. limicola*, *Chlorobium limicola*; *O. granulosus*, *Oceanicola granulosus*; *P. horikoshii*, *Pyrococcus horikoshii*; *T. denitrificans*, *Thiobacillus denitrificans*.

a drastic effect on substrate binding and activity of the resultant mutant proteins (32). Based on mutant analysis, N123 was ascribed a role in the catalytic steps that follow the enolization of substrate RuBP. It is replaced by E119 in the *C. tepidum* RLP and K98 in YkrW. E119 in *C. tepidum* RLP still seems to have the ability to form hydrogen bonds with the backbone of the substrate (Fig. 10). As discussed above, K98 in YkrW is thought to abstract protons and serve as the general base (33). K177 of form I RubisCO appears to participate in catalysis by controlling the pK_a of K175, which acts as both a proton

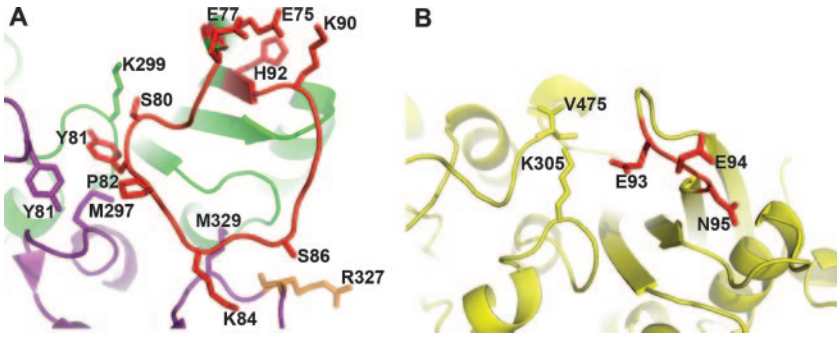


FIG. 14. Comparison of the unique loop CD of *C. tepidum* RLP (PDB accession number 1YKW) (A) with the comparable region of form I (spinach) RubisCO (accession number 8RUC) (B). Residues Q78 to I91 form a loop (loop CD) (red ribbon and sticks), and residues in this loop have multiple interactions with residues of the same subunit (green ribbon and sticks) or the neighboring large subunit (purple ribbon and sticks). Notably, the hydroxyl group of S86 forms a hydrogen bond with loop 6 residue R327 (orange sticks) from the neighboring large subunit. Spinach form I residues equivalent to E75, E77, and H92 of *C. tepidum* RLP are E93, E94, and N95 (red ribbon and sticks). Residues K305 and V475 (yellow sticks) interact with E93 in the closed conformation of spinach RubisCO.

acceptor and donor at two different steps in the catalytic mechanism (13). K177 is replaced by N174 in *C. tepidum* RLP and V152/M149 in YkrW, with the side chains of these residues being more distant from the substrate binding site. K334 of form I RubisCO is at the apex of flexible loop 6 that folds over the active site and controls the CO_2/O_2 specificity of RubisCO. Dynamically, loop 6 of RubisCO, which is in an open conformation, is thought to close upon substrate binding, bringing K334 closer to the substrate and thus forming a hydrogen bond with the incoming carboxyl group during carboxylation of RuBP (20). Because of the importance of K334 in the catalytic mechanism, loop 6 has been extensively studied with various approaches. Amino acid substitutions/modifications at K334 or at any of the residues in the vicinity appear to have drastic effects on catalysis (66). As noted above, K334 is analogous to R327 of the *C. tepidum* RLP and S305 of YkrW (33, 39). A K334R substitution in *R. rubrum* RubisCO led to a complete loss of activity, suggesting the criticality of the chemical nature of this residue for RubisCO function (64). Thus, the analogous R327, whose side chains appear in two different conformations in *C. tepidum* RLP and *R. palustris* RLP2 (Fig. 10), and other residues in the vicinity may play a crucial role in the catalytic reaction mechanism(s) of various RLPs.

Comparison of Secondary Structural Elements Unique to RLP and RubisCO: Possible Implications for RLP Structure-Function Relationships

As noted above, sequence and structural alignments of the three bona fide forms of RubisCO with the form IV RLPs (Fig. 12 and 13) indicate that there are at least two regions in the secondary structure of RLPs that differ from the bona fide RubisCO enzymes. A loop comprised of at least five or more residues connecting β-sheets C and D (loop CD) appears to be absent in the structures of RubisCO (Fig. 12). Loop CD is comprised of residues Q78 to I91 in *C. tepidum* RLP. These residues are involved in multiple interactions close to the active site (Fig. 14) and hence may be critical for the function of RLP in vivo. In the structure of YkrW, loop CD becomes a helix and forms a stronger interaction interface with the other monomer of the dimer. Such interactions are unique to RLP/YkrW structures, and hence, variations in the lengths and residue identities of loop CD may confer differences in properties among various RLPs (Fig. 12). The analogous regions in plant and algal form I RubisCO enzymes are involved with interactions with a class of proteins known as RubisCO activases (49). RubisCO activase is a member of the family of AAA^+ proteins (ATPases associated with a variety of cellular activities) characterized by chaperonin-like functions. Activases interact with RubisCO in an ATP-dependent manner to release tight-binding sugar phosphates from the active sites prior to catalysis. They are found in all plants and green algae, and an activase-like gene has also been identified in filamentous cyanobacteria (*Anabaena* and related species) (40). Drawing a parallel, one may argue that the CD loops in RLPs could

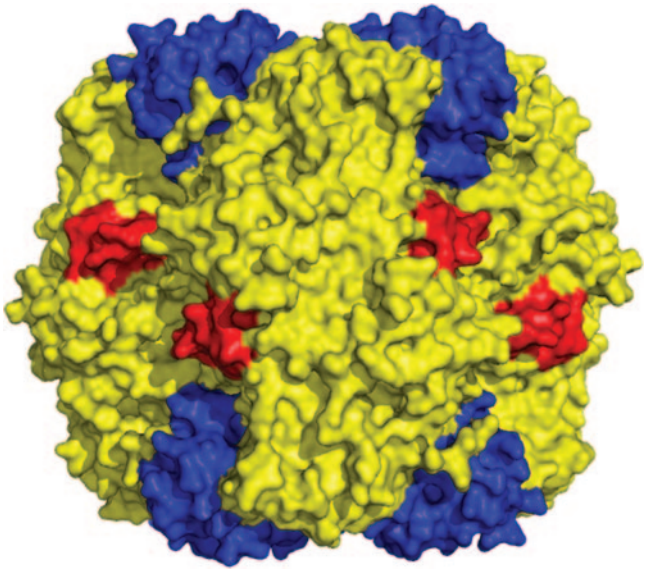


FIG. 15. Placement of the β-hairpin residues in the holoenzyme structure of form I (spinach) RubisCO (PDB accession number 8RUC). The β-hairpin residues (Y353 to S367) (red) that are absent in RLPs are exposed to the solvent in the holoenzyme structure of spinach RubisCO. The large subunits are yellow, and the small subunits are blue.

potentially act as a regulatory structural element gating the active sites.

The second structural region that demarcates RLPs from the three forms of RubisCO is a β-hairpin structure that appears to be juxtaposed by the N-terminal domain on one side and the C-terminal domain on the other side in all three forms of bona fide authentic RubisCO enzymes (Fig. 15). The strategic placement of these elements in RubisCO indicates that this secondary structural element may mediate conformational changes and maintain the relative positions of the N- and C-terminal domains. Although none of the residues in this region appear to be involved in critical interactions with the active site of RubisCO, the side chains of most of these residues are polar in nature and are solvent exposed in the holoenzyme (Fig. 15). The absence of β-hairpin structures in RLPs may account for the differences in structural stabilities between bona fide RubisCO enzymes and RLPs. Although the *C. tepidum* RLP functions as a dimer, the *T. kodakarensis* form III RubisCO (decamer or pentamer of dimers) and the spinach form I RubisCO (L_8S_8 hexadecamer) appear to be more closely related to the *C. tepidum* RLP than the dimeric *R. rubrum* form II RubisCO based on structural analyses (39). Attempts to decipher the functional relationships in RLPs via genetic engineering strategies targeting individual amino acid residues as well as secondary structural elements must also consider the implications of such changes on the gross alteration of the holoenzyme structure. A combination of tools such as DNA shuffling, random mutagenesis, and bioselection may be exploited to delineate the physiological role of RLPs.

CONCLUSIONS AND OUTLOOK

About 30 years have passed since it was discovered that microbes synthesize RubisCO molecules that differ from the typical plant paradigm. Clearly, three separate bona fide forms of RubisCO (forms I, II, and III) have now been described, each of which catalyzes the carboxylation or oxygenation of RuBP, albeit for potentially different physiological purposes. Moreover, a fourth class, the RLPs, or form IV proteins, is clearly structurally related to bona fide RubisCO, yet the RLPs do not function as RubisCO enzymes, but thus far, they all seem to catalyze reactions involved in sulfur metabolism. However, RubisCO and some RLPs do possess functional similarities in that both proteins catalyze reactions using analogous substrates in both cases via an initial enolization-type reaction. The great preponderance of RLP sequences now available has further shown that there are, at present, six different clades of RLPs, some of which appear to possess different physiological roles. Indeed, RubisCO and RLP molecules have now been described for each of the three recognized types of living organisms, and the huge number of sequences now available has allowed a coherent picture of the likely evolutional events that took place to account for the different classes of RubisCOs and RLPs to emerge. Our analyses are compatible with an archaeal origin of both RubisCO and RLP, with form III proteins from the *Methanomicrobia* being the likely precursors for all modern RubisCO and RLP lineages. Certainly, as additional information becomes available, we and others will build upon and/or challenge this hypothesis. However, at this time, no other evolutional scheme is compatible with the data. Finally, structural and functional studies of RubisCO and RLP will continue to provide information as to how the active sites of these proteins have become adapted for their specific functions.

ACKNOWLEDGMENTS

Work in our laboratories was supported by NIH grant GM24497 and DOE grant DE-FG02-91ER20033 (F.R.T.), by NSF Career Award MCB-0447649 (T.E.H.), and by DOE-BER (H.L. and S.C.).

We thank Simona Romagnoli for the *rlp2* clone and Yim Wu for her assistance in protein purification and crystallization.

REFERENCES

1. **Abascal, F., R. Zardoya, and D. Posada.** 2005. Prottest: selection of best-fit models of protein evolution. Bioinformatics **21:**2104–2105.
2. **Alfreider, A., C. Vogt, D. Hoffmann, and W. Babel.** 2003. Diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes from groundwater and aquifer microorganisms. Microb. Ecol. **45:**317–328.
3. **Anantharaman, V., L. Aravind, and E. V. Koonin.** 2003. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. Curr. Opin. Chem. Biol. **7:**12–20.
4. **Andersson, I.** 1996. Large structures at high resolution: the 1.6 Å crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. J. Mol. Biol. **259:**160–174.
5. **Andersson, I., and T. C. Taylor.** 2003. Structural framework for catalysis and regulation in ribulose-1,5-bisphosphate carboxylase/oxygenase. Arch. Biochem. Biophys. **414:**130–140.
6. **Andreeva, A., D. Howorth, S. E. Brenner, T. J. Hubbard, C. Chothia, and A. G. Murzin.** 2004. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res. **32:**D226–229.
7. **Ashida, H., A. Danchin, and A. Yokota.** 2005. Was photosynthetic Rubisco recruited by acquisitive evolution from Rubisco-like proteins involved in sulfur metabolism? Res. Microbiol. **156:**611–618.
8. **Ashida, H., Y. Saito, C. Kojima, K. Kobayashi, N. Ogasawara, and A. Yokota.** 2003. A functional link between Rubisco-like protein of *Bacillus* and photosynthetic Rubisco. Science **302:**286–290.
9. **Beatty, J. T., J. Overmann, M. T. Lince, A. K. Manske, A. S. Lang, R. E. Blankenship, C. L. Van Dover, T. A. Martinson, and F. G. Plumley.** 2005. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proc. Natl. Acad. Sci. USA **102:**9306–9310.
10. **Bowers, P. M., M. Pellegrini, M. J. Thompson, J. Fierro, T. O. Yeates, and D. Eisenberg.** 2004. Prolinks: a database of protein functional linkages derived from coevolution. Genome Biol. **5:**R35.
11. **Carre-Mlouka, A., A. Mejean, P. Quillardet, H. Ashida, Y. Saito, A. Yokota, I. Callebaut, A. Sekowska, E. Dittmann, C. Bouchier, and N. T. de Marsac.** 2006. A new Rubisco-like protein coexists with a photosynthetic Rubisco in the planktonic cyanobacteria *Microcystis*. J. Biol. Chem. **281:**24462–24471.
12. **Chan, L. K., R. Morgan-Kiss, and T. E. Hanson.** Sulfur oxidation in *Chlorobium tepidum* (syn. *Chlorobaculum tepidum*): genetic and proteomic analyses. *In* C. Dahl and C. G. Friedrich (ed.), Proceedings of the International Symposium on Microbial Sulfur Metabolism, in press. Springer, New York, NY.
13. **Cleland, W. W., J. T. Andrews, S. Gutteridge, F. C. Hartman, and G. H. Lorimer.** 1998. Mechanism of RubisCO: the carbamate as general base. Chem. Rev. **98:**549–561.
14. **Dandekar, T., B. Snel, M. Huynen, and P. Bork.** 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem. Sci. **23:**324–328.
15. **DeBry, R. W.** 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. Mol. Biol. Evol. **9:**537–551.
16. **Delwiche, C. F., and J. D. Palmer.** 1996. Rampant horizontal transfer and duplication of Rubisco genes in eubacteria and plastids. Mol. Biol. Evol. **13:**873–882.
17. **Derelle, E., C. Ferraz, S. Rombauts, P. Rouze, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynie, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piegu, S. G. Ball, J. P. Ral, F. Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, and H. Moreau.** 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. USA **103:**11647–11652.
18. **Doolittle, W. F.** 1999. Phylogenetic classification and the universal tree. Science **284:**2124–2129.
19. **Dubbs, J. M., and F. R. Tabita.** 2004. Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO_2 assimilation, nitrogen fixation, hydrogen metabolism and energy generation. FEMS Microbiol. Rev. **28:**353–376.
20. **Duff, A. P., T. J. Andrews, and P. M. Curmi.** 2000. The transition between the open and closed states of Rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. J. Mol. Biol. **298:**903–916.

21. **Ellis, R. J.** 1979. Most abundant protein in the world. Trends Biochem. Sci. **4:**241–244.
22. **Elsaied, H., and T. Naganuma.** 2001. Phylogenetic diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes from deep-sea microorganisms. Appl. Environ. Microbiol. **67:**1751–1765.
23. **Elsaied, H. E., H. Kimura, and T. Naganuma.** 2007. Composition of archaeal, bacterial, and eukaryal RuBisCO genotypes in three Western Pacific arc hydrothermal vent systems. Extremophiles **11:**191–202.
24. **Enright, A. J., I. Iliopoulos, N. C. Kyrpides, and C. A. Ouzounis.** 1999. Protein interaction maps for complete genomes based on gene fusion events. Nature **402:**86–90.
25. **Ezaki, S., N. Maeda, T. Kishimoto, H. Atomi, and T. Imanaka.** 1999. Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, Pyrococcus kodakaraensis KOD1. J. Biol. Chem. **274:**5078–5082.
26. **Finn, M. W., and F. R. Tabita.** 2004. Modified pathway to synthesize ribulose 1,5-bisphosphate in methanogenic archaea. J. Bacteriol. **186:**6360–6366.
27. **Finn, M. W., and F. R. Tabita.** 2003. Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. J. Bacteriol. **185:**3049–3059.
28. **Fuchs, G., S. Lange, E. Rude, S. Schaefer, R. Schauder, R. Scholtz, and E. Stupperich.** 1987. Autotrophic CO_2 fixation in chemotrophic anaerobic bacteria, p. 39–43. *In* H. W. Van Verseveld and J. A. Duine (ed.), Microbial growth on C_1 compounds. Martinus Nijhoff, Dordrecht, The Netherlands.
29. **Gibson, J. L., and F. R. Tabita.** 1977. Different molecular forms of D-ribulose-1,5-bisphosphate carboxylase from *Rhodopseudomonas sphaeroides*. J. Biol. Chem. **252:**943–949.
30. **Hanson, T. E., and F. R. Tabita.** 2003. Insights into the stress response and sulfur metabolism revealed by proteome analysis of a *Chlorobium tepidum* mutant lacking the Rubisco-like protein. Photosynth. Res. **78:** 231–248.
31. **Hanson, T. E., and F. R. Tabita.** 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. Proc. Natl. Acad. Sci. USA **98:**4397–4402.
32. **Hartman, F. C., and M. R. Harpel.** 1994. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Annu. Rev. Biochem. **63:**197–234.
33. **Imker, H. J., A. A. Fedorov, E. V. Fedorov, S. C. Almo, and J. A. Gerlt.** 2007. Mechanistic diversity in the RuBisCO superfamily: the "enolase" in the methionine salvage pathway in *Geobacillus kaustophilus*. Biochemistry **46:** 4077–4089.
34. **John, D. E., B. Wawrik, F. Tabita, and J. H. Paul.** 2006. Gene diversity and organization in *rbcL*-containing genome fragments from uncultivated *Synechococcus* in the Gulf of Mexico. Mar. Ecol. Prog. Ser. **316:**23–33.
35. **Knight, S., I. Andersson, and C. I. Branden.** 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. Subunit interactions and active site. J. Mol. Biol. **215:**113–160.
36. **Kopp, J., S. Kopriva, K. H. Suss, and G. E. Schulz.** 1999. Structure and mechanism of the amphibolic enzyme D-ribulose-5-phosphate 3-epimerase from potato chloroplasts. J. Mol. Biol. **287:**761–771.
37. **Kreel, N. E., and F. R. Tabita.** 2007. Substitutions at methionine 295 of *Archaeoglobus fulgidus* ribulose-1,5-bisphosphate carboxylase/oxygenase affect oxygen binding and CO_2/O_2 specificity. J. Biol. Chem. **282:**1341–1351.
38. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
39. **Li, H., M. R. Sawaya, F. R. Tabita, and D. Eisenberg.** 2005. Crystal structure of a RuBisCO-like protein from the green sulfur bacterium *Chlorobium tepidum*. Structure **13:**779–789.
40. **Li, L. A., and F. R. Tabita.** 1994. Transcription control of ribulose bisphosphate carboxylase/oxygenase activase and adjacent genes in *Anabaena* species. J. Bacteriol. **176:**6697–6706.
41. **Lo, I., V. J. Denef, N. C. Verberkmoes, M. B. Shah, D. Goltsman, G. DiBartolo, G. W. Tyson, E. E. Allen, R. J. Ram, J. C. Detter, P. Richardson, M. P. Thelen, R. L. Hettich, and J. F. Banfield.** 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature **446:**537–541.
42. **Lundqvist, T., and G. Schneider.** 1991. Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II), and activator CO_2 at 2.3-Å resolution. Biochemistry **30:**904–908.
43. **Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg.** 1999. Detecting protein function and protein-protein interactions from genome sequences. Science **285:**751–753.
44. **Morse, D., P. Salois, P. Markovic, and J. W. Hastings.** 1995. A nuclear-encoded form II RuBisCO in dinoflagellates. Science **268:**1622–1624.
45. **Murphy, B. A., F. J. Grundy, and T. M. Henkin.** 2002. Prediction of gene function in methylthioadenosine recycling from regulatory signals. J. Bacteriol. **184:**2314–2318.
46. **Nagano, N., C. A. Orengo, and J. M. Thornton.** 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. **321:**741–765.
47. **Nanba, K., G. M. King, and K. Dunfield.** 2004. Analysis of facultative lithotroph distribution and diversity on volcanic deposits by use of the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase. Appl. Environ. Microbiol. **70:**2245–2253.
48. **Novotny, M., D. Madsen, and G. J. Kleywegt.** 2004. Evaluation of protein fold comparison servers. Proteins **54:**260–270.
49. **Ott, C. M., B. D. Smith, A. R. Portis, Jr., and R. J. Spreitzer.** 2000. Activase region on chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase. Nonconservative substitution in the large subunit alters species specificity of protein interaction. J. Biol. Chem. **275:**26241–26244.
50. **Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev.** 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA **96:**2896–2901.
51. **Palmer, J. D.** 1995. Rubisco rules fall; gene transfer triumphs. Bioessays **17:**1005–1008.
52. **Palmer, J. D.** 1996. Rubisco surprises in dinoflagellates. Plant Cell **8:**343–345.
53. **Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates.** 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc. Natl. Acad. Sci. USA **96:**4285–4288.
54. **Pellegrini, M., M. Thompson, J. Fierro, and P. Bowers.** 2001. Computational method to assign microbial genes to pathways. J. Cell. Biochem. Suppl. **37:**106–109.
55. **Pichard, S. L., L. Campbell, and J. H. Paul.** 1997. Diversity of the ribulose bisphosphate carboxylase/oxygenase form I gene (*rbcL*) in natural phytoplankton communities. Appl. Environ. Microbiol. **63:**3600–3606.
56. **Rhee, S., K. D. Parris, C. C. Hyde, S. A. Ahmed, E. W. Miles, and D. R. Davies.** 1997. Crystal structures of a mutant (betaK87T) tryptophan synthase alpha2beta2 complex with ligands bound to the active sites of the alpha- and beta-subunits reveal ligand-induced conformational changes. Biochemistry **36:**7664–7680.
57. **Robbens, S., E. Derelle, C. Ferraz, J. Wuyts, H. Moreau, and Y. Van de Peer.** 2007. The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. Mol. Biol. Evol. **24:**956–968.
58. **Rowan, R., S. M. Whitney, A. Fowler, and D. Yellowlees.** 1996. Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. Plant Cell **8:**539–553.
59. **Sato, T., H. Atomi, and T. Imanaka.** 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science **315:**1003–1006.
60. **Saunders, N. F., T. Thomas, P. M. Curmi, J. S. Mattick, E. Kuczek, R. Slade, J. Davis, P. D. Franzmann, D. Boone, K. Rusterholtz, R. Feldman, C. Gates, S. Bench, K. Sowers, K. Kadner, A. Aerts, P. Dehal, C. Detter, T. Glavina, S. Lucas, P. Richardson, F. Larimer, L. Hauser, M. Land, and R. Cavicchioli.** 2003. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. Genome Res. **13:**1580–1588.
61. **Schneider, G., Y. Lindqvist, and C. I. Branden.** 1992. Rubisco: structure and mechanism. Annu. Rev. Biophys. Biomol. Struct. **21:**119–143.
62. **Schreuder, H. A., S. Knight, P. M. Curmi, I. Andersson, D. Cascio, C. I. Branden, and D. Eisenberg.** 1993. Formation of the active site of ribulose-1,5-bisphosphate carboxylase/oxygenase by a disorder-order transition from the unactivated to the activated form. Proc. Natl. Acad. Sci. USA **90:**9968–9972.
63. **Sekowska, A., and A. Danchin.** 2002. The methionine salvage pathway in *Bacillus subtilis*. BMC Microbiol. **2:**8.
64. **Soper, T. S., R. J. Mural, F. W. Larimer, E. H. Lee, R. Machanoff, and F. C. Hartman.** 1988. Essentiality of Lys-329 of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* as demonstrated by site-directed mutagenesis. Protein Eng. **2:**39–44.
65. **Spiridonova, E. M., I. A. Berg, T. V. Kolganova, R. N. Ivanovskii, B. B. Kuznetsov, and T. P. Turova.** 2004. An oligonucleotide primer system for amplification of the ribulose-1,5-bisphosphate carboxylase/oxygenase genes of bacteria of various taxonomic groups. Mikrobiologiia **73:**377–387. (In Russian.)
66. **Spreitzer, R. J., and M. E. Salvucci.** 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. Annu. Rev. Plant Biol. **53:**449–475.
67. **Tabita, F. R.** 1995. The biochemistry and regulation of carbon metabolism and CO_2 fixation in purple bacteria, p. 885–914. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), Anoxygenic photosynthetic bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
68. **Tabita, F. R.** 1999. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. Photosynth. Res. **60:**1–28.
69. **Tabita, F. R., and B. A. McFadden.** 1974. D-Ribulose 1,5-diphosphate carboxylase from *Rhodospirillum rubrum*. I. Levels, purification, and effects of metallic ions. J. Biol. Chem. **249:**3453–3458.
70. **Tabita, F. R., and B. A. McFadden.** 1974. D-Ribulose 1,5-diphosphate car-

boxylase from *Rhodospirillum rubrum*. II. Quaternary structure, composition, catalytic, and immunological properties. J. Biol. Chem. **249:**3459–3464.
71. Reference deleted.
72. **Taylor, T. C., A. Backlund, K. Bjorhall, R. J. Spreitzer, and I. Andersson.** 2001. First crystal structure of Rubisco from a green alga, Chlamydomonas reinhardtii. J. Biol. Chem. **276:**48159–48164.
73. **Watson, G. M., J. P. Yu, and F. R. Tabita.** 1999. Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic archaea. J. Bacteriol. **181:** 1569–1575.
74. **Watson, G. M. F., and F. R. Tabita.** 1997. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a molecule for phylogenetic and enzymological investigation. FEMS Microbiol. Lett. **146:**13–22.
75. **Wawrik, B., J. H. Paul, and F. R. Tabita.** 2002. Real-time PCR quantification of *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl. Environ. Microbiol. **68:**3771–3779.
76. **Wierenga, R. K.** 2001. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett. **492:**193–198.
77. **Wise, E., W. S. Yew, P. C. Babbitt, J. A. Gerlt, and I. Rayment.** 2002. Homologous (beta/alpha)-barrel enzymes that catalyze unrelated reactions: orotidine 5′-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase. Biochemistry **41:**3861–3869.
78. **Xu, H. H., and F. R. Tabita.** 1996. Ribulose-1,5-bisphosphate carboxylase/oxygenase gene expression and diversity of Lake Erie planktonic microorganisms. Appl. Environ. Microbiol. **62:**1913–1921.
79. **Yoon, K. S., T. E. Hanson, J. L. Gibson, and F. R. Tabita.** 2000. Autotrophic CO_2 metabolism, p. 349–358. *In* J. Lederburg (ed.), Encyclopedia of microbiology, 2nd ed. Academic Press Inc., San Diego, CA.
80. **Yooseph, S., G. Sutton, D. B. Rusch, A. L. Halpern, S. J. Williamson, K. Remington, J. A. Eisen, K. B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C. S. Miller, H. Li, S. T. Mashiyama, M. P. Joachimiak, C. van Belle, J. M. Chandonia, D. A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B. J. Raphael, V. Bafna, R. Friedman, S. E. Brenner, A. Godzik, D. Eisenberg, J. E. Dixon, S. S. Taylor, R. L. Strausberg, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. PLoS Biol. **5:**e16.